

VISIT...

LANZAROTE
Caliente.COM

В.М. ВЕРЖБИЦКИЙ

ЧИСЛЕННЫЕ МЕТОДЫ

Математический
анализ
и обыкновенные
дифференциальные
уравнения



ВЫСШАЯ ШКОЛА

В.М. ВЕРЖБИЦКИЙ

ЧИСЛЕННЫЕ МЕТОДЫ

**Математический
анализ
и обыкновенные
дифференциальные
уравнения**

**Допущено
Министерством образования
Российской Федерации
в качестве учебного пособия
для студентов высших учебных заведений,
обучающихся по математическим
специальностям и направлениям подготовки
дипломированных специалистов в области
техники и технологии**



**Москва
«Высшая школа» 2001**

УДК 519.6
ББК 22.19
В31

Рецензенты :

кафедра математического моделирования систем и процессов
Пермского государственного технического университета (зав. кафедрой —
д-р физ.-мат. наук, проф. П.В.Трусов); кафедра вычислительной матема-
тики Удмуртского государственного университета (зав. кафедрой —
д-р физ.-мат. наук, проф. Г.Г.Исламов); д-р физ.-мат. наук, проф.
Н.В.Азбелев

ISBN 5-06-003982-X

© ГУП «Издательство «Высшая школа», 2001

Оригинал-макет данного издания является собственностью издатель-
ства «Высшая школа» и его репродуцирование (воспроизведение) любым
способом без согласия издательства запрещено.

ОГЛАВЛЕНИЕ

Предисловие	6
Глава 1. Полиномиальная интерполяция	
§ 1.1. Задача и способы аппроксимации функций	8
§ 1.2. Интерполяционный многочлен Лагранжа	11
§ 1.3. Интерполяционная схема Эйткена	20
§ 1.4. Конечные разности	26
§ 1.5. Конечноразностные интерполяционные формулы	32
§ 1.6. Интерполяционная формула Ньютона для неравноотстоящих узлов	45
§ 1.7. Обратное интерполирование	51
§ 1.8. Интерполяция с кратными узлами	56
Упражнения	61
Глава 2. Многочлены Чебышева и наилучшие равномерные приближения	
§ 2.1. Определение и свойства многочленов Чебышева	64
§ 2.2. Интерполяция по чебышевским узлам	69
§ 2.3. О многочленах наилучших равномерных приближений	72
§ 2.4. Экономизация степенных рядов	78
Упражнения	82
Глава 3. Метод наименьших квадратов и наилучшие среднеквадратические приближения	
§ 3.1. Простейшая обработка эмпирических данных методом наименьших квадратов	84
§ 3.2. Обобщенные многочлены наилучших среднеквадратических приближений	92
§ 3.3. О нормальной системе МНК при полиномиальной аппроксимации	96
§ 3.4. Системы ортогональных многочленов	100
§ 3.5. Простая процедура построения системы ортогональных многочленов	103
§ 3.6. Аппроксимация функций многочленами Фурье	106
Упражнения	109
Глава 4. Интерполяционные сплайны	
§ 4.1. Кусочно-полиномиальная аппроксимация. Линейные фильтры	111
§ 4.2. Определение сплайна. Интерполяционный кубический сплайн дефекта 1	117
§ 4.3. Квадратичный сплайн дефекта 1	125
§ 4.4. Базисные сплайны	133

§ 4.5. Эрмитовы (локальные) сплайны	138
Упражнения	144

Глава 5. Численное интегрирование

§ 5.1. Задача численного интегрирования. Квадратурные формулы прямоугольников	145
§ 5.2. Семейство квадратурных формул Ньютона–Котеса	151
§ 5.3. Составные квадратурные формулы трапеций и Симпсона	158
§ 5.4. Соотношения между формулами прямоугольников, трапеций и Симпсона	161
§ 5.5. Принцип Рунге практического оценивания погрешностей. Алгоритм Ромберга	163
§ 5.6. Квадратурные формулы Чебышева и Гаусса	167
§ 5.7. Формулы Гаусса–Кристоффеля	175
§ 5.8. Приемы приближенного вычисления несобственных интегралов	181
Упражнения	187

Глава 6. Аппроксимация производных

§ 6.1. Вывод формул численного дифференцирования	189
§ 6.2. Остаточные члены простейших формул численного дифференцирования	193
§ 6.3. Оптимизация шага численного дифференцирования при ограниченной точности значений функции	203
Упражнения	208

Глава 7. Методы Эйлера и Рунге–Кутты решения начальных задач для обыкновенных дифференциальных уравнений

§ 7.1. Постановка задачи. Классификация приближенных методов. Метод последовательных приближений	210
§ 7.2. Метод Эйлера — разные подходы к построению	214
§ 7.3. Несколько простых модификаций метода Эйлера	218
§ 7.4. Исправленный метод Эйлера	222
§ 7.5. О семействе методов Рунге–Кутты. Методы второго порядка	223
§ 7.6. Методы Рунге–Кутты произвольного и четвертого порядков	225
§ 7.7. Пошаговый контроль точности. Метод Кутты–Мерсона	228
Упражнения	232

Глава 8. Линейные многошаговые методы

§ 8.1. Многошаговые методы Адамса	234
§ 8.2. Методы прогноза и коррекции. Предиктор-корректорные методы Адамса	241
§ 8.3. Метод Милна четвертого порядка	244

§ 8.4. Общий вид линейных многошаговых методов. Условия согласованности	247
§ 8.5. О численном решении систем дифференциальных уравнений первого порядка	253
§ 8.6. Численное решение дифференциальных уравнений высших порядков. Методы Адамса–Штёрмера	255
Упражнения	260
Глава 9. О проблемах численной устойчивости	
§ 9.1. Общая схема решения задач численного анализа. Аппроксимация, устойчивость, сходимость	262
§ 9.2. Простейшие разностные аппроксимации задачи Коши. Глобальная погрешность метода Эйлера	266
§ 9.3. Краткие сведения о решениях линейных разностных уравнений с постоянными коэффициентами	270
§ 9.4. Устойчивость и неустойчивость некоторых простейших разностных схем	273
§ 9.5. Исследование устойчивости многошаговых методов ..	277
§ 9.6. Жесткие уравнения и системы	280
§ 9.7 A - и $A(\alpha)$ -устойчивость. Чисто неявные методы	285
Упражнения	292
Глава 10. Методы приближенного решения краевых задач для обыкновенных дифференциальных уравнений	
§ 10.1. Постановка задачи. Классификация приближенных методов	294
§ 10.2. Методы сведения краевых задач к начальным	296
§ 10.3. Метод конечных разностей	302
§ 10.4. Метод коллокации	308
§ 10.5. Метод Галёркина	314
§ 10.6. Метод конечных элементов	319
Упражнения	330
Глава 11. Численное решение интегральных уравнений	
§ 11.1. Некоторые общие сведения об интегральных уравнениях	333
§ 11.2. Квадратурный метод решения интегральных уравнений Фредгольма	341
§ 11.3. Квадратурный метод решения интегральных уравнений Вольтерра	345
§ 11.4. Квадратурно-итерационный метод построения резольвент	355
Упражнения	362
Приложение. Образцы постановок лабораторных заданий	364
Литература	372
Предметный указатель	378

ПРЕДИСЛОВИЕ

Эта книга является продолжением предыдущей книги автора [Вержбицкий В.М. Численные методы (линейная алгебра и нелинейные уравнения): Учеб. пособие для вузов. — М.: Высш. шк., 2000] и имеет одинаковую с ней структуру. Весь включенный в книгу материал разбит на 11 глав, каждая из которых начинается с аннотации и заканчивается упражнениями. Нумерация параграфов и формул — своя в пределах главы, номер которой отражен в первой позиции (до разделяющей точки) обозначений параграфов и формул. Это относится также к нумерации определений, утверждений разных типов, примеров, упражнений, рисунков и таблиц. Как и в предыдущей книге, здесь имеется приложение, содержащее постановки заданий для лабораторных работ, без выполнения которых автору представляется трудным рассчитывать на успешность освоения изучаемых численных методов. Разумеется, предполагается, что результаты лабораторных работ должны сопоставляться и тщательно анализироваться. В списке литературы наряду с учебниками, монографиями, справочниками указаны некоторые задачки по численным методам (см. [10, 33, 45, 66, 74]), с помощью которых можно расширить спектр приводимых здесь примеров, упражнений и лабораторных работ.

Учитывая, что классические методы численного анализа, как правило, опираются на замену функций близкими им многочленами, вопросам приближения функций многочленами (полиномиальной аппроксимации в широком смысле) посвящены первые четыре главы. При этом наиболее подробно рассматривается полиномиальная интерполяция (глава 1), на основе которой далее выводятся формулы численного интегрирования (глава 5) и дифференцирования (глава 6). Использование многочленов Чебышева (глава 2) позволяет минимизировать погрешность интерполяционной формулы Лагранжа, а также получить некоторые квадратурные формулы наряду с другими такими формулами, базирующимися на ортогональных многочленах (глава 3). Сплайны, изучаемые в главе 4, представляют самостоятельный интерес как современный инструмент для описания многочленами невысоких степеней функций, заданных таблицами своих значений, а B -сплайны первого порядка используются далее в главе 10 при выводе метода конечных элементов решения линейной краевой задачи для обыкновенного дифференциального уравнения второго порядка. Главы 7 и 8 посвящены соответственно одношаговым и многошаговым методам решения задач Коши с широким привлечением формул интерполирования, квадратурных формул и формул аппроксимации производных разностными отношениями; реальное поведение разных численных процессов решения начальных задач, в том числе задач, называемых жесткими, исследуется в главе 9. К главам 7–10, где изучаются дифференциальные уравнения, примыкает глава 11, в которой показывается, как можно численно находить решения

линейных интегральных уравнений и их резольвент, привлекая квадратурные формулы главы 5.

Так же, как и в предыдущей книге, наряду со знаком « $=$ » здесь часто используется знак « $:=$ », который, в зависимости от контекста, интерпретируется двояко:

- 1) *присвоить, положить равным;*
- 2) *считать по определению, ввести величину, обозначить.*

Изредка употребляется и знак « $\approx$ », означающий *считать приближенно равным.*

Из аббревиатур, которые могут встретиться в тексте, отметим три:

СЛАУ — *система линейных алгебраических уравнений,*

МНК — *метод наименьших квадратов,*

ОДУ — *обыкновенное дифференциальное уравнение.*

По умолчанию переменные i, j, k, l, m, n принимают только целые неотрицательные значения, т.е. принадлежат множеству N_0 или какому-то его подмножеству; по записи переменной « x » или « x » можно сразу определить, идет ли речь о векторной или о скалярной величине. Вводимые в книгу термины (которые далеко не всегда специально строго определены, но все их можно найти в предметном указателе) выделяются жирным курсивом, важные части текста (также не всегда оформленные в виде четких утверждений) — простым курсивом.

Невзирая на тесную связь этой книги с упомянутой предыдущей, содержащийся в ней материал может изучаться совершенно независимо. При этом отметим важность умения эффективно решать линейные алгебраические системы той или иной структуры, чтобы некоторые из рассматриваемых здесь методов могли быть успешно доведены до конца; в частности, речь идет о методе прогонки решения СЛАУ с трехдиагональными матрицами коэффициентов, который требуется применять в нескольких случаях.

Автор выражает глубокую благодарность своим коллегам по кафедре «Прикладная математика и информатика» Ижевского государственного технического университета доценту А.А.Айзиковичу за содействие в оформлении рукописи и профессору А.Л.Тептину за внимательное ее прочтение и сделанные замечания. Автор признателен всем, кто принял участие в компьютерном наборе рукописи: О.Батуеву, С.Высоцкому, А.Дерюшеву, В.Матвееву, М.Петрову (рисунки), Д.Хохрякову, Д.Чепелю, А.Чуракову и особенно Ю.Гаврину, придавшему набранной не без его наставлений рукописи окончательный вид. Автор также благодарит профессоров Н.В.Азбелева, Г.Г.Исламова и П.В.Трусова, взявшим на себя нелегкий труд по рецензированию рукописи.

Автор

ГЛАВА 1 || ПОЛИНОМИАЛЬНАЯ || ИНТЕРПОЛЯЦИЯ

Прежде всего, обсуждаются различные постановки задачи аппроксимации функций одной переменной, в частности, полиномиальной аппроксимации. Далее конкретизируется задача интерполяции, определяется решающий эту задачу интерполяционный многочлен Лагранжа, доказывается его единственность, выводится остаточный член. Рассматривается итерационный принцип вычисления промежуточных значений таблично заданных (сеточных) функций с помощью лагранжевой интерполяции, известный как интерполяционная схема Эйткена. Для случая равноотстоящих узлов вводятся конечные разности, отмечаются их простейшие свойства, строятся конечно-разностные интерполяционные многочлены, в структуру которых заложено убывание значимости каждого последующего слагаемого. Формулы такой же структуры, использующие аппарат разделенных разностей, выводятся и для неравных промежутков между узлами. Завершается глава изучением задачи обратного интерполирования и задачи интерполирования с кратными узлами; последняя объединяет в себе лагранжеву интерполяцию и локальную аппроксимацию функций по формуле Тейлора.

§ 1.1. ЗАДАЧА И СПОСОБЫ АППРОКСИМАЦИИ ФУНКЦИЙ

В основе большинства численных методов математического анализа лежит подмена одной функции $f(x)$ (известной, неизвестной или частично известной) другой функцией $\varphi(x)$, близкой к $f(x)$ и обладающей «хорошими» свойствами, позволяющими легко производить над нею те или иные аналитические или вычислительные операции. Будем называть такую подмену *аппроксимацией*^{*)} или просто *приближением* функции $f(x)$ функцией $\varphi(x)$. Для того, чтобы построить какую-то разумную теорию таких приближений и предложить конкретные способы получения аппроксимирующих функций $\varphi(x)$ по заданным тем

^{*)} Часто термин «аппроксимация функций» используется в более узком смысле, чем это принимается здесь.

или иным образом аппроксимируемыми функциям $f(x)$, предварительно следует ответить на ряд вопросов.

А именно:

1) *Что известно о функции $f(x)$?* Задана ли она своим аналитическим выражением или таблицей своих значений, какова степень ее гладкости и доступны ли значения ее производных, как расположены точки в интересующей части области определения $f(x)$, где известны ее значения, и можно ли их задавать по своему усмотрению, и т.п.

2) *Какому классу (семейству) функций должна принадлежать функция $\varphi(x)$?* Какие дополнительные требования предъявляются к $\varphi(x)$, выделяющие ее из заданного класса?

3) *Что понимать под близостью между $f(x)$ и $\varphi(x)$; иначе, какой принять **критерий согласия** между ними?* Говоря языком функционального анализа, по метрике какого пространства должно быть малым расстояние между $f(x)$ и $\varphi(x)$?

Как видим, задача аппроксимации функции $f(x)$ функцией $\varphi(x)$ состоит в построении для заданной функции $f(x)$ такой функции $\varphi(x)$, что

$$f(x) \approx \varphi(x), \quad (1.1)$$

причем левая часть приближенного равенства (1.1) должна быть обусловлена ответами на вопросы первой группы, правая часть — второй группы, а ответ на вопрос 3) должен уточнить значение связывающего $f(x)$ и $\varphi(x)$ символа $\approx$.

Прежде всего, определимся с ответом на второй вопрос. *Договоримся использовать в качестве аппроксимирующих функций $\varphi(x)$ только многочлены или функции, составленные из многочленов**; в таком случае будем говорить о **полиномиальной аппроксимации** или **кусочно-полиномиальной аппроксимации** соответственно.

По сравнению с другими семействами функций, пригодных для построения теории приближений, например, таких, как тригонометрические или показательные функции, для вычислительной математики многочлены привлекательны тем, что они являются линейными функциями своих параметров (коэффициентов), и их вычисление сводится к выполнению конечного числа простейших арифметических операций — сложения и умножения.

Будем считать, что аппроксимация функции $f(x)$ произво-

* За небольшим исключением в § 3.1.

дится с помощью многочленов степени $n \in \mathbb{N}_0$. Тогда в зависимости от выбора критерия согласия и, в частности, от количества точек согласования $f(x)$ с $\varphi(x)$ (будем называть их узлами), т.е. точек, в которых известна информация об $f(x)$ и, возможно, ее производных, можно рассмотреть разные конкретные способы аппроксимации. Некоторые из них, нашедшие отражение в табл.1.1, будут изучаться ниже. А именно, в данной главе будет достаточно подробно рассматриваться классическая лагранжева интерполяция (клетка 1.2 в табл.1.1), служащая основой многих численных методов, в частности, приближенного дифференцирования и интегрирования. Здесь же можно будет получить представление об интерполяционном многочлене Эрмита (клетка 2.1), частными случаями которого являются, с одной стороны, многочлен Тейлора (при $i = 1$), с другой — многочлен Лагранжа (при $i = n + 1$). Понятие о многочленах наилучших равномерных приближений (клетка 3.3) дается в главе 2, наилучшие среднеквадратические приближения (клетка 4.3) освещаются в главе 3, а кусочно-полиномиальной аппроксимации (клетка 1.3) и сплайнам (клетка 2.3) посвящена глава 4.

Таблица 1.1

Тип функции $\varphi(x)$, аппроксимирующей $f(x)$, выраженной через многочлены степени n (при разных условиях согласования)

Количество точек согласования (узлов) Критерий согласия	$(1 \leq i \leq n + 1)$	$n + 1$	$m (\geq n + 1)$
Совпадение значений функций $f(x)$ и $\varphi(x)$ в узлах	1.1	1.2 Интерполяционный многочлен Лагранжа	1.3 Кусочно-полиномиальная функция
Совпадение в узлах функций $f(x)$ и $\varphi(x)$ и некоторых их производных	2.1 Интерполяционный многочлен Эрмита	2.2	2.3 Интерполяционный сплайн
Минимум максимального отклонения (на отрезке)	3.1	3.2	3.3 Многочлен наилучшего равномерного приближения
Минимум среднеквадратического отклонения (в узлах или на отрезке)	4.1	4.2	4.3 Многочлен наилучшего среднеквадратического приближения

§ 1.2. ИНТЕРПОЛЯЦИОННЫЙ МНОГОЧЛЕН ЛАГРАНЖА

Пусть в точках $x_0, x_1, \dots, x_n$ таких, что $a \leq x_0 < \dots < x_n \leq b$, известны значения функции $y = f(x)$, т.е. на отрезке $[a, b]$ задана **табличная (сеточная) функция**

$$f(x): \begin{array}{c|c|c|c|c} x & x_0 & x_1 & \dots & x_n \\ \hline y & y_0 & y_1 & \dots & y_n \end{array}. \quad (1.2)$$

Функция $\varphi(x)$ называется **интерполирующей** или **интерполяционной** для $f(x)$ на $[a, b]$, если ее значения $\varphi(x_0), \varphi(x_1), \dots, \varphi(x_n)$ в заданных точках $x_0, x_1, \dots, x_n$, называемых **узлами интерполяции**, совпадают с заданными значениями функции $f(x)$, т.е. с $y_0, y_1, \dots, y_n$ соответственно.*¹⁾ Геометрически факт интерполирования означает, что график функции $\varphi(x)$ проходит так, что, по меньшей мере, в $n + 1$ заданных точках он пересекает или касается графика функции $f(x)$ (рис.1.1).

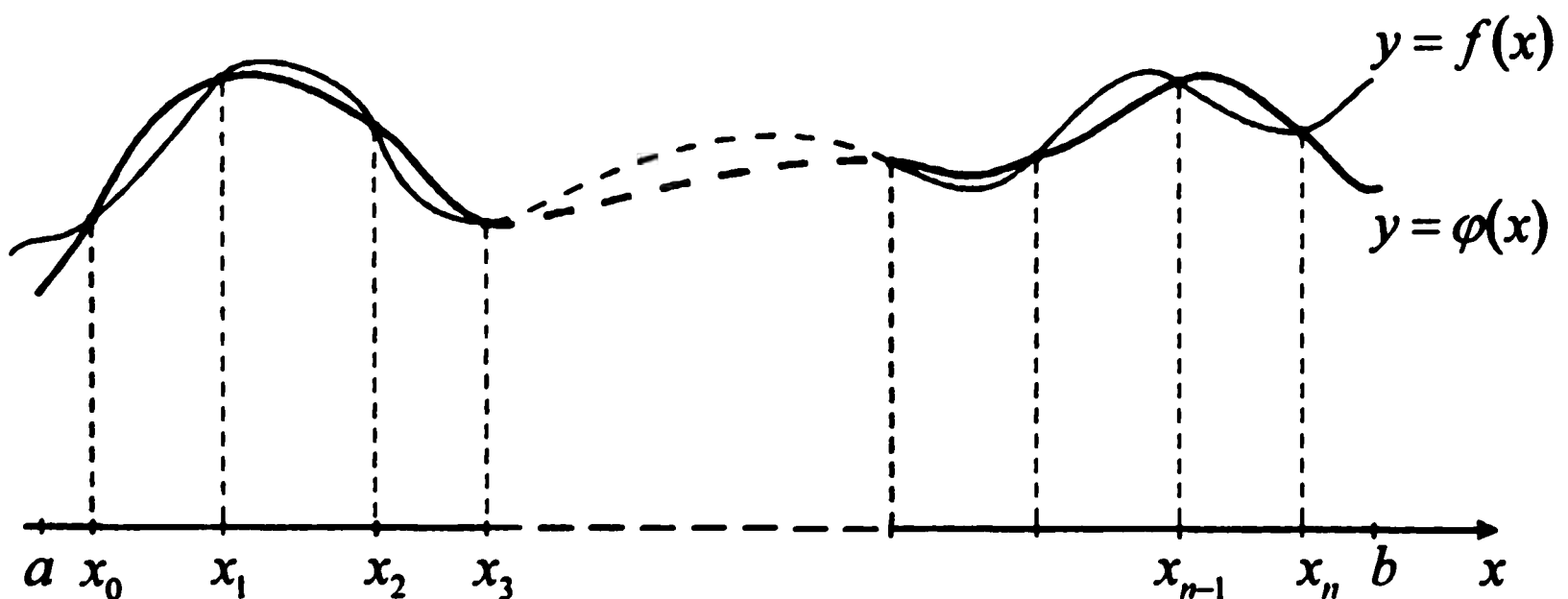


Рис.1.1. Геометрическая интерпретация задачи интерполирования

Легко представить, что таких графиков $\varphi(x)$, проходящих через заданные точки, можно изобразить сколько угодно, и они могут отличаться от графика $f(x)$ сколь угодно сильно, если не накладывать на $\varphi(x)$ и $f(x)$ определенных ограничений.

*¹⁾ Латинское слово *interpolatio* переводится как обновление, изменение, переделка. Обычно термин **интерполяция** (или иначе, **интерполирование**) означает процесс построения интерполяционной функции или процесс нахождения промежуточных значений табличной функции. Этот термин введен в 1656 году английским математиком Джоном Валлисом (Уоллисом, 1616–1703 гг.); кстати отметим, что годом раньше он ввел ныне общепринятый символ бесконечности (∞) [96].

потребовать выполнения условия

$$l_i(x_j) = \delta_{ij} := \begin{cases} 0, & \text{если } j \neq i \\ 1, & \text{если } j = i \end{cases} \quad \forall j, i \in \{0, 1, \dots, n\}. \quad (1.5)$$

В таком случае для многочлена

$$L_n(x) = \sum_{i=0}^n y_i l_i(x)$$

в каждом узле x_j ($j \in \{0, 1, \dots, n\}$), в силу (1.5), справедливо

$$\begin{aligned} L_n(x_j) &= l_0(x_j)y_0 + \dots + l_{j-1}(x_j)y_{j-1} + l_j(x_j)y_j + l_{j+1}(x_j)y_{j+1} + \dots \\ &\quad \dots + l_n(x_j)y_n = 0 + \dots + 0 + y_j + 0 + \dots + 0 = y_j, \end{aligned}$$

т.е. выполняются условия интерполяции (1.3).

Чтобы конкретизировать базисные многочлены $l_i(x)$, учтем, что они должны удовлетворять условиям (1.5). Равенство нулю i -го многочлена во всех узлах, кроме i -го, означает, что $l_i(x)$ можно записать в виде

$$l_i(x) = A_i(x - x_0) \dots (x - x_{i-1})(x - x_{i+1}) \dots (x - x_n),$$

а коэффициент A_i этого представления легко получается из содержащегося в (1.5) требования $l_i(x_i) = 1$. Подставляя в выражение $l_i(x)$ значение $x = x_i$ и приравнявая результат единице, получаем

$$A_i = \frac{1}{(x_i - x_0) \dots (x_i - x_{i-1})(x_i - x_{i+1}) \dots (x_i - x_n)}.$$

Таким образом, **базисные многочлены Лагранжа** имеют вид

$$l_i(x) = \frac{(x - x_0)(x - x_1) \dots (x - x_{i-1})(x - x_{i+1}) \dots (x - x_n)}{(x_i - x_0)(x_i - x_1) \dots (x_i - x_{i-1})(x_i - x_{i+1}) \dots (x_i - x_n)},$$

а искомый **интерполяционный многочлен Лагранжа** есть

$$L_n(x) = \sum_{i=0}^n \frac{(x - x_0) \dots (x - x_{i-1})(x - x_{i+1}) \dots (x - x_n)}{(x_i - x_0) \dots (x_i - x_{i-1})(x_i - x_{i+1}) \dots (x_i - x_n)} y_i. \quad (1.6)$$

Заметим, что числитель, фигурирующий в записи i -го слагаемого $L_n(x)$ дроби, представляет собой произведение разностей между переменной x и всеми узлами, кроме i -го, а знаменатель — произведение разностей между i -м узлом и всеми остальными.

В качестве примера запишем интерполяционные многочлены Лагранжа первой и второй степени.

При $n=1$ информация об интерполируемой функции $y = f(x)$ сосредоточена в двух точках: $(x_0; y_0)$ и $(x_1; y_1)$. Многочлен Лагранжа в этом случае составляется с помощью двух базисных многочленов первой степени ($l_0(x)$ и $l_1(x)$) и имеет вид

$$L_1(x) = \frac{x - x_1}{x_0 - x_1} y_0 + \frac{x - x_0}{x_1 - x_0} y_1. \quad (1.7)$$

При $n=2$ по трехточечной таблице

$$f(x): \begin{array}{c|c|c|c} x & x_0 & x_1 & x_2 \\ \hline y & y_0 & y_1 & y_2 \end{array}$$

можно образовать три базисных многочлена ($l_0(x)$, $l_1(x)$ и $l_2(x)$) и, соответственно, интерполяционный многочлен Лагранжа второй степени

$$L_2(x) = \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} y_0 + \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} y_1 + \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)} y_2. \quad (1.8)$$

Приближенные равенства

$$f(x) \approx L_1(x) \quad \text{и} \quad f(x) \approx L_2(x)$$

называют соответственно **формулами линейной и квадратичной интерполяции**. Геометрически они означают подмену графика функции $y = f(x)$ на некотором отрезке $[a, b]$ оси абсцисс, содержащем точки x_0, x_1 в первом и x_0, x_1, x_2 во втором случаях, соответствующими участками прямой линии и квадратичной параболы, проходящих через заданные точки координатной плоскости.

Такая простейшая интерполяция широко применялась при составлении различных таблиц значений функций для их пополнения промежуточными значениями (что и являлось основной задачей интерполяции на ранней стадии развития вычислительной математики). В связи с этим, заметим, что иногда термину **интерполяция** противопоставляется термин **экстраполяция**. В таких случаях речь идет о том, что под интерполяцией понимается нахождение промежуточных значений таблично заданной

функции строго внутри таблицы, тогда как экстраполяция^{*)} предполагает использование интерполяционного многочлена, построенного по значениям функции $f(x)$ в точках $x_0, x_1, \dots, x_n$, для нахождения ее приближенных значений за пределами промежутка $[x_0, x_n]$.

Вернемся к изучению интерполяционного многочлена Лагранжа (1.6).

Покажем его **единственность** (от противного). Предположим, что наряду с $L_n(x)$ имеется другой многочлен n -й степени $Q_n(x)$, решающий ту же задачу интерполяции, т.е. удовлетворяющий условиям интерполяции типа (1.3):

$$Q_n(x_i) = y_i \quad \forall i \in \{0, 1, \dots, n\}.$$

Образует новый многочлен как разность между $L_n(x)$ и $Q_n(x)$. Этот многочлен $P_n(x) := L_n(x) - Q_n(x)$ имеет степень не выше n и во всех $n+1$ узлах $x_0, x_1, \dots, x_n$ обращается в нуль, в силу равенства значений $Q_n(x_i)$ и $L_n(x_i)$ одним и тем же числам y_i . Получается, что точки $x_0, x_1, \dots, x_n$ служат корнями многочлена $P_n(x)$. Но по следствию из основной теоремы алгебры многочленов $P_n(x)$ не может иметь более n корней. Полученное противоречие означает, что многочлены $Q_n(x)$ и $L_n(x)$ должны полностью совпадать, т.е. по заданным $n+1$ значению функции можно построить единственный интерполяционный многочлен.

Пусть для данной функции $f(x)$ интерполяционный многочлен $L_n(x)$ построен, т.е. для приближенного представления функции $f(x)$ на отрезке $[a, b] \supseteq [x_0, x_n]$ применяется **интерполяционная формула**

$$f(x) \approx L_n(x). \quad (1.9)$$

Естественно встает вопрос: какова погрешность такого приближенного равенства? Иначе, сколь велико может быть различие между значениями интерполируемой функции $f(x)$ и соответствующими значениями интерполяционного многочлена Лагранжа $L_n(x)$ в точках отрезка $[a, b]$, не совпадающих с узловыми точками?

^{*)} Указанный здесь узкий смысл терминов «интерполирование» и «экстраполирование» становится очевидным, если учитывать их латинское происхождение: «inter» и «extra» означают соответственно «между» и «вне», а «polire» — «делать гладким» [97]; таким образом, интерполирование — это сглаживание между узлами, а экстраполирование — сглаживание вне таблицы.

Для совершенно произвольной функции $f(x)$ такая постановка вопроса о погрешности интерполяции заведомо некорректна, поскольку функций, для которых построенный (единственный!) многочлен $L_n(x)$ будет интерполяционным, бесконечно много; легко представить себе эту ситуацию графически: через заданные $n+1$ точку с координатами $(x_0; y_0), (x_1; y_1), \dots, (x_n; y_n)$, согласно доказанному, можно провести единственную параболу — график многочлена степени n — и, в то же время, можно изобразить сколько угодно графиков других функций, как угодно сильно отличающихся от этой параболы. Этот факт говорит о том, что загнать величину этого отклонения, т.е. погрешность интерполяции, в определенные рамки невозможно, если не наложить каких-то ограничений на гладкость интерполируемой функции $f(x)$ и на расположение узлов интерполяции $x_0, x_1, \dots, x_n$ на отрезке $[a, b]$.

Будем выяснять величину отклонения $f(x)$ от $L_n(x)$ в произвольной точке $x \in [a, b]$, иначе, величину *остаточного члена*

$$R_n(x) := f(x) - L_n(x)$$

интерполяционной формулы Лагранжа (1.9) в предположении, что $f(x) \in C_{[a, b]}^{n+1}$, т.е. данная функция $n+1$ раз непрерывно дифференцируема.

Обозначим

$$\Pi_{n+1}(x) := \prod_{i=0}^n (x - x_i) = (x - x_0)(x - x_1) \dots (x - x_n) \quad (1.10)$$

— определенный через узлы $x_0, x_1, \dots, x_n$ многочлен $(n+1)$ -й степени. Через него введем в рассмотрение функцию

$$u(x) := f(x) - L_n(x) - c\Pi_{n+1}(x), \quad (1.11)$$

где c — некоторая постоянная (параметр).

Так как в точках $x = x_0, x_1, \dots, x_n$ многочлен $\Pi_{n+1}(x)$ обращается в нуль, согласно его конструкции, а $f(x) - L_n(x) = 0$ в этих точках по условиям интерполяции, то и $u(x_i) = 0$ при $i = 0, 1, \dots, n$, т.е. функция $u(x)$ имеет на отрезке $[a, b]$ по меньшей мере $n+1$ корень. Подберем параметр c так, чтобы $u(x)$ имела заведомо еще и $(n+2)$ -й корень в какой-то фиксированной точке $\bar{x} (\neq x_i)$ промежутка $[a, b]$. Имеем:

$$u(\bar{x}) = 0 \Leftrightarrow f(\bar{x}) - L_n(\bar{x}) = c\Pi_{n+1}(\bar{x}) \Leftrightarrow c = \frac{f(\bar{x}) - L_n(\bar{x})}{\Pi_{n+1}(\bar{x})},$$

причем такое значение c обязательно найдется, поскольку $\Pi_{n+1}(x) = 0$ только в узлах x_i .

Пусть для определенности $\bar{x} \in (x_i, x_{i+1})$. Тогда можно утверждать, что при найденном c функция $u(x)$ равна нулю на концах $n+1$ отрезков $[x_0, x_1], [x_1, x_2], \dots, [x_i, \bar{x}], [\bar{x}, x_{i+1}], \dots, [x_{n-1}, x_n]$. Значит, к функции $u(x)$ на каждом из этих отрезков применима теорема Ролля, т.е. внутри каждого из этих отрезков существует, по крайней мере, по одной такой точке, в которой производная функции $u(x)$ обращается в нуль. Эти $n+1$ точки образуют систему из n отрезков, на концах каждого из которых уже функция $u'(x)$ равна нулю, т.е. теперь к производной можно применить теорему Ролля, по которой существует n нулей второй производной функции $u(x)$. Продолжая процесс таких рассуждений далее, в конце концов, приходим к выводу о существовании такой точки $\xi \in (x_0, x_n) \subseteq (a, b)$, что $u^{(n+1)}(\xi) = 0$. Учитывая, что n -я производная многочлена n -й степени постоянна, а $(n+1)$ -я равна нулю, находим выражение $(n+1)$ -й производной функции $u(x)$, заданной равенством (1.11):

$$u^{(n+1)}(x) = f^{(n+1)}(x) - 0 - c(n+1)!.$$

Итак, существует точка $\xi \in (x_0, x_n)$ такая, что

$$f^{(n+1)}(\xi) - c(n+1)! = 0, \quad \text{т.е.} \quad c = \frac{f^{(n+1)}(\xi)}{(n+1)!}.$$

Это значение c должно совпадать с выбранным ранее, т.е. должно выполняться равенство

$$\frac{f(\bar{x}) - L_n(\bar{x})}{\Pi_{n+1}(\bar{x})} = \frac{f^{(n+1)}(\xi)}{(n+1)!},$$

откуда получаем

$$f(\bar{x}) - L_n(\bar{x}) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \Pi_{n+1}(\bar{x}).$$

Так как в качестве $\bar{x}$ могла быть взята любая точка x из промежутка $[a, b]$, не совпадающая ни с какой узловой, расфиксируем (или, как еще говорят, разморозим) точку $\bar{x}$, т.е. заменим ее в последнем равенстве произвольной точкой $x \neq x_i$, в результате чего приходим к выражению остаточного члена

$$R_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \Pi_{n+1}(x). \quad (1.12)$$

Знание остаточного члена в предположении $(n+1)$ -кратной дифференцируемости $f(x)$ позволяет записать точное представление $f(x)$ через ее интерполяционный многочлен $L_n(x)$:

$$f(x) = L_n(x) + \frac{f^{(n+1)}(\xi)}{(n+1)!} \Pi_{n+1}(x), \quad (1.13)$$

где ξ — некоторая (вообще говоря, неизвестная, причем зависящая от x) точка из промежутка интерполяции (a, b) , а $\Pi_{n+1}(x)$ — определенный в (1.10) многочлен^{*}.

Теперь можно ставить и пытаться отвечать на вопросы о погрешности приближенного вычисления значения $f(x)$ с помощью $L_n(x)$ в какой-либо конкретной точке из $[a, b]$, о величине максимальной погрешности, допускаемой при подмене функции $f(x)$ многочленом $L_n(x)$ на отрезке $[a, b]$, о сходимости интерполяционного процесса, т.е. о том, имеет ли место $\rho(f(x), L_n(x)) \xrightarrow{n \rightarrow \infty} 0$ по метрике $\rho(\cdot, \cdot)$ того или иного определенного на $[a, b]$ функционального пространства.

Так, если известна величина

$$M_{n+1} := \max_{x \in [a, b]} |f^{(n+1)}(x)|,$$

то оценить абсолютную погрешность интерполяционной формулы (1.9) в любой точке $\tilde{x} \in [a, b]$ можно с помощью неравенства

$$|R_n(\tilde{x})| = |f(\tilde{x}) - L_n(\tilde{x})| \leq \frac{M_{n+1}}{(n+1)!} |\Pi_{n+1}(\tilde{x})|. \quad (1.14)$$

Максимальная погрешность интерполирования на отрезке $[a, b]$ оценивается величиной

$$\max_{x \in [a, b]} |R_n(x)| \leq \frac{M_{n+1}}{(n+1)!} \max_{x \in [a, b]} |\Pi_{n+1}(x)|. \quad (1.15)^q$$

Так как максимумы функций $f^{(n+1)}(x)$ и $\Pi_{n+1}(x)$ достигаются, вообще говоря, в разных точках отрезка $[a, b]$, то более точной, но более трудно реализуемой по сравнению с (1.15) следует считать оценку

$$\max_{x \in [a, b]} |R_n(x)| \leq \frac{1}{(n+1)!} \max_{x \in [a, b]} |f^{(n+1)}(x) \Pi_{n+1}(x)|.$$

^{*} Формула (1.12) устанавливалась для значений $x \neq x_j$. Но при $x = x_j$ левая и правая части (1.12) равны нулю, следовательно, формула (1.12), а с нею и представление (1.13), справедливы при любых $x \in [a, b]$.

Пример 1.1. Рассмотрим квадратичную интерполяцию функции $y = \sin x$ на отрезке $\left[0, \frac{\pi}{2}\right]$ по ее трем значениям: $\sin 0 = 0$, $\sin \frac{\pi}{4} = \frac{\sqrt{2}}{2}$, $\sin \frac{\pi}{2} = 1$.

По формуле (1.8) строим многочлен Лагранжа второй степени

$$L_2(x) = \frac{\left(x - \frac{\pi}{4}\right)\left(x - \frac{\pi}{2}\right)}{\left(0 - \frac{\pi}{4}\right)\left(0 - \frac{\pi}{2}\right)} \cdot 0 + \frac{(x - 0)\left(x - \frac{\pi}{2}\right)}{\left(\frac{\pi}{4} - 0\right)\left(\frac{\pi}{4} - \frac{\pi}{2}\right)} \cdot \frac{\sqrt{2}}{2} + \frac{(x - 0)\left(x - \frac{\pi}{4}\right)}{\left(\frac{\pi}{2} - 0\right)\left(\frac{\pi}{2} - \frac{\pi}{4}\right)} \cdot 1,$$

или после преобразований

$$L_2(x) = \frac{8}{\pi^2} x \left[(1 - \sqrt{2})x + \left(\frac{\sqrt{2}}{2} - \frac{1}{4} \right) \pi \right]. \quad (1.16)$$

Остаточный член для этого случая получаем по формуле (1.12), учитывая, что $n = 2$, $(\sin x)''' = -\cos x$ и $\Pi_3(x) = x \left(x - \frac{\pi}{4} \right) \left(x - \frac{\pi}{2} \right)$.

Имеем:

$$R_2(x) = \sin x - L_2(x) = \frac{-\cos \xi}{3!} x \left(x - \frac{\pi}{4} \right) \left(x - \frac{\pi}{2} \right).$$

Так как точка $\xi \in \left(0, \frac{\pi}{2}\right)$ неизвестна, можно делать лишь оценки $|R_2(x)|$, полагая $M_3 = \max_{x \in \left[0, \frac{\pi}{2}\right]} |-\cos x| = 1$. Найдя максимальное значение $|\Pi_3(x)|$,

реализуемое в двух точках данного отрезка $x_{1,2} = \frac{\pi}{12} (3 \pm \sqrt{3})$ и не превосходящее 0.568, по формуле (1.15) оцениваем сверху величину допустимого отклонения дуги параболы (1.16) от данной синусоиды на промежутке интерполирования:

$$\max_{x \in \left[0, \frac{\pi}{2}\right]} |\sin x - L_2(x)| \leq \frac{1}{3!} \max_{x \in \left[0, \frac{\pi}{2}\right]} |\Pi_3(x)| < \frac{0.568}{6} \approx 0.095.$$

Подставим в полученный интерполяционный многочлен (1.16) контрольную точку $\tilde{x} = \frac{\pi}{6}$. Получим приближенное значение

$$L_2\left(\frac{\pi}{6}\right) = \frac{4\sqrt{2} - 1}{9} \approx 0.517, \text{ отличающееся от значения } \sin \frac{\pi}{6} = 0.5 \text{ на вели-}$$

чину ≈ 0.017 , меньшую, чем это допускается оценкой по формуле (1.14):

$$\left| \sin \frac{\pi}{6} - L_2 \left(\frac{\pi}{6} \right) \right| \leq \frac{M_3}{3!} \left| \Pi_3 \left(\frac{\pi}{6} \right) \right| = \frac{1}{6} \left(\frac{\pi}{6} \right)^3 \approx 0.075.$$

Наблюдаем типичную картину: фактическая погрешность меньше оценки погрешности в точке, которая, в свою очередь, меньше максимальной погрешности на отрезке (т.е. по чебышевской норме).

В заключение этого параграфа заметим, что через введенный в (1.10) многочлен $\Pi_{n+1}(x)$ интерполяционный многочлен Лагранжа (1.6) можно записать в более компактной форме. Для этого достаточно увидеть, что знаменатель фигурирующей там дроби представляет собой значение производной многочлена $\Pi_{n+1}(x)$ в i -м узле, а числитель есть просто $\Pi_{n+1}(x)$ без множителя $x - x_i$.

Таким образом,

$$L_n(x) = \sum_{i=0}^n \frac{\Pi_{n+1}(x)y_i}{(x - x_i)\Pi'_{n+1}(x_i)} = \Pi_{n+1}(x) \sum_{i=0}^n \frac{y_i}{(x - x_i)\Pi'_{n+1}(x_i)}. \quad (1.6')$$

Вопросы сходимости $L_n(x)$ к $f(x)$ при $n \rightarrow \infty$ будут обсуждаться несколько позже (см. § 1.5.).

§ 1.3. ИНТЕРПОЛЯЦИОННАЯ СХЕМА ЭЙТКЕНА

Пусть функция $f(x)$ и расположение узлов $x_0, x_1, \dots, x_n$ на промежутке интерполяции $[a, b]$ таковы, что имеет место сходимость процесса интерполяции, т.е. $R_n(x) \rightarrow 0$ при $n \rightarrow \infty$, и пусть решается частная задача вычисления отдельных приближенных значений функции $f(x)$ с помощью вычисления соответствующих им значений интерполяционного многочлена Лагранжа $L_n(x)$. Для построения эффективного способа решения такой частной задачи интерполяции примем во внимание следующие три обстоятельства.

Во-первых, непосредственное использование многочлена Лагранжа в форме (1.6) неудобно из-за его громоздкости (что чревато большими вычислительными затратами). Во-вторых, как правило, заранее неизвестно, какой степени многочлен нужно использовать для интерполирования данной функции с требуемой точностью, а постепенное наращивание точности за счет повторных вычислений значений $L_n(x)$ со все большими показателями степени n (подобно тому, как это можно делать, например, при вычислении значений функции по формуле Тейлора) при прямом применении формулы (1.6) малопримлемо, в силу пло-

хой перестраиваемости $L_{n-1}(x)$ в $L_n(x)$. В-третьих, при реальном счете всегда следует помнить, что функция $f(x)$ задается таблицей своих приближенных значений, и каноническое развитие процесса сходимости $L_n(x)$ к $f(x)$ при больших значениях n будет нарушено все возрастающим влиянием на результат исходных ошибок (более подробно об этом см. далее в §§ 1.4, 1.5).

Построим вычислительную схему для получения приближенного значения сеточной функции $f(x)$ в заданной точке $x = \tilde{x}$, в основу которой будет положена лагранжева интерполяция на сетке узлов $x_0, x_1, \dots, x_n$ и организация вычислений по которой будет иметь итерационный характер. Каждый итерационный шаг в этой схеме заключается в вычислении некоторого определителя второго порядка.

Пусть даны две точки на кривой $y = f(x)$: $(x_0; y_0)$ и $(x_1; y_1)$. Введем функцию $P_{0,1}(x)$ через определитель следующим образом:

$$P_{0,1}(x) := \frac{1}{x_1 - x_0} \begin{vmatrix} x - x_0 & y_0 \\ x - x_1 & y_1 \end{vmatrix}.$$

Раскрыв этот определитель, видим, что

$$P_{0,1}(x) = \frac{x - x_1}{x_0 - x_1} y_0 + \frac{x - x_0}{x_1 - x_0} y_1 = L_1(x),$$

т.е. $P_{0,1}(x)$ совпадает с интерполяционным многочленом Лагранжа первой степени, построенным по данным двум точкам (ср. с (1.7)).

Если взять на кривой $y = f(x)$ точки $(x_1; y_1)$ и $(x_2; y_2)$, то, очевидно, функция

$$P_{1,2}(x) := \frac{1}{x_2 - x_1} \begin{vmatrix} x - x_1 & y_1 \\ x - x_2 & y_2 \end{vmatrix},$$

в развернутой форме имеющая вид

$$P_{1,2}(x) = \frac{x - x_2}{x_1 - x_2} y_1 + \frac{x - x_1}{x_2 - x_1} y_2,$$

тоже является многочленом Лагранжа первой степени, интерполирующим $f(x)$ по точкам $(x_1; y_1)$ и $(x_2; y_2)$.

Считая, что на кривой $y = f(x)$ заданы три точки $(x_0; y_0)$, $(x_1; y_1)$ и $(x_2; y_2)$, с помощью введенных линейных функций $P_{0,1}(x)$ и $P_{1,2}(x)$ образуем новую функцию

$$P_{0,1,2}(x) := \frac{1}{x_2 - x_0} \begin{vmatrix} x - x_0 & P_{0,1}(x) \\ x - x_2 & P_{1,2}(x) \end{vmatrix}.$$

Легко видеть, что эта функция есть многочлен второй степени (точнее, не выше второй). Учитывая, что

$$P_{0,1}(x_0) = P_{1,2}(x_1) = y_0, \quad P_{0,1}(x_1) = y_1 \quad \text{и} \quad P_{1,2}(x_2) = y_2,$$

подстановкой в $P_{0,1,2}(x)$ поочередно значений $x = x_0, x_1, x_2$ убеждаемся, что $P_{0,1,2}(x_0) = y_0$, $P_{0,1,2}(x_1) = y_1$, $P_{0,1,2}(x_2) = y_2$. Таким образом, функция $P_{0,1,2}(x)$ есть многочлен второй степени, решающий задачу параболической интерполяции по трем точкам $(x_0; y_0)$, $(x_1; y_1)$, $(x_2; y_2)$. Но такой многочлен, как доказано в предыдущем параграфе, единствен, следовательно, $P_{0,1,2}(x) = L_2(x)$, где $L_2(x)$ — многочлен Лагранжа (1.8).

Продолжая так рассуждать и далее, приходим к рекуррентному заданию последовательности интерполяционных многочленов Лагранжа, которое и составляет суть так называемой *интерполяционной схемы Эйткена*:

$$f(x) \approx P_{0,1,\dots,i}(x) = \frac{1}{x_i - x_0} \begin{vmatrix} x - x_0 & P_{0,1,\dots,i-1}(x) \\ x - x_i & P_{1,2,\dots,i}(x) \end{vmatrix}, \quad (1.17)$$

где $i = 1, 2, \dots, n$ и, по определению, $P_0(x) := y_0$, $P_1(x) := y_1$.

Методом математической индукции можно показать идентичность (1.17) при $i = n$ и многочлена Лагранжа (1.6).

Пример 1.2. Пусть некоторая функция $y = y(x)$ задана таблицей своих значений, округленных до двух знаков после запятой:

x_i	0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4
y_i	0.00	0.45	0.63	0.77	0.89	1.00	1.10	1.18

Рассмотрим поведение процесса вычисления двух значений этой функции по схеме Эйткена в точках: а) $\tilde{x} = 0.1$; б) $\tilde{x} = 0.5$. Результаты промежуточных вычислений (в которых один знак — запасной) сведем в табл. 1.2 и 1.3 для случаев а) и б) соответственно. Числа в столбцах, помеченных посредством $P_{\overline{i,i+k}}(\tilde{x})$, представляют собой значения многочленов Лагранжа k -й степени, построенных по узлам от i -го до $(i+k)$ -го рекуррентно по формуле

$$P_{\overline{i,i+k}}(\tilde{x}) = \frac{1}{x_{i+k} - x_i} \begin{vmatrix} \tilde{x} - x_i & P_{\overline{i,i+k-1}}(\tilde{x}) \\ \tilde{x} - x_{i+k} & P_{\overline{i+1,i+k}}(\tilde{x}) \end{vmatrix}, \quad (1.18)$$

где $k = 1, 2, \dots$; $P_{\overline{i,i}} := y_i$, в соответствии с интерполяционной схемой Эйткена (1.17).

Таблица 1.2

Последовательность значений интерполяционных множителей, участвующих в схеме Зйткена при вычислении значения $y(0.1)$

i	x_i	y_i	$P_{i,j+1}(0.1)$	$P_{i,j+2}(0.1)$	$P_{i,j+3}(0.1)$	$P_{i,j+4}(0.1)$	$P_{i,j+5}(0.1)$	$P_{i,j+6}(0.1)$	$P_{i,j+7}(0.1)$
0	0	0.00	10.225	30.259	60.273	100.281	150.287	210.291	280.246
1	0.2	0.45	20.360	50.345	90.339	140.336	200.336	270.336	
2	0.4	0.63	40.420	80.383	130.362	190.338	260.339		
3	0.6	0.77	70.470	120.426	180.426	250.336			
4	0.8	0.89	110.505	170.426	240.570				
5	1	1.00	160.550	230.303					
6	1.2	1.10	220.660						
7	1.4	1.18							

$$y(0.1) \approx P_{0,6}(0.1) \approx 0.29$$

Таблица 1.3

Вычисление по схеме Эйткена значения $y(0.5)$

i	x_i	y_i	$P_{\overline{i,i+1}}(0.5)$	$P_{\overline{i,i+2}}(0.5)$	$P_{\overline{i,i+3}}(0.5)$	$P_{\overline{i,i+4}}(0.5)$	$P_{\overline{i,i+5}}(0.5)$
0	0	0.00					
1	0.2	0.45					
2	0.4	0.63	10.700	30.700			
3	0.6	0.77	20.700				
4	0.8	0.89					
5	1	1.00					
6	1.2	1.10		$y(0.5) \approx P_{\overline{2,3}}(0.5) \approx P_{\overline{2,4}}(0.5) \approx 0.70$			
7	1.4	1.18					

Организация вычислений по формуле (1.18) должна быть такова, что если заранее неизвестна степень интерполяционного многочлена, который следует использовать для вычисления $y(\tilde{x})$, и данная таблица значений функции достаточно обширна, то должно происходить постепенное повышение степени k интерполирующих ее многочленов за счет подключения новых, все более удаленных от $\tilde{x}$ узлов. Порядок заполнения клеток табл.1.2 и 1.3 получаемыми по формуле (1.18) числами показан проставленными в этих клетках номерами. Счет ведется до тех пор, пока идет уточнение приближенного значения $y(\tilde{x})$, о чем можно судить по уменьшению величины $\left|P_{\overline{i,i+k-1}}(\tilde{x}) - P_{\overline{i,i+k}}(\tilde{x})\right|$ при увеличении k и подходящем фиксировании i .

Так, для подсчета приближенного значения данной функции в точке $\tilde{x} = 0.1$, расположенной между узлами $x_0 = 0$ и $x_1 = 0.2$, целесообразно в качестве основной последовательности значений интерполяционных многочленов Лагранжа брать числа $P_{\overline{0,1}}(0.1) = 0.225 = L_1(0.1)$, $P_{\overline{0,2}}(0.1) = 0.259 = L_2(0.1)$ и т.д., т.е. строку табл.1.2, соответствующую значению $i = 0$. Составив разности между последующими и предыдущими числами этой строки, а именно:

$$0.034 \quad 0.014 \quad 0.008 \quad 0.006 \quad 0.004 \quad -0.045,$$

видим, что дальнейший счет бессмыслен; нарушается каноническое развитие процесса итерационного уточнения, и по данной информации о функции $y(x)$ более точное значение $y(0.1)$, чем значение $y(0.1) \approx 0.29$,

получаемое с помощью $L_6(0.1)$, найти не удастся^{*)}.

В случае б) для вычисления значения $y(0.5)$ оказалось достаточным сделать лишь линейную интерполяцию по двум ближайшим к точке $\tilde{x} = 0.5$ узлам $x_2 = 0.4$ и $x_3 = 0.6$ и с помощью квадратичной интерполяции по узлам $x_2 = 0.4$, $x_3 = 0.6$ и $x_4 = 0.8$ убедиться, что полученное линейной интерполяцией значение $y(0.1) \approx 0.70$ при этом не изменилось (см. табл.1.3).

Столь большое различие в степенях интерполяционных многочленов и, соответственно, в объеме вычислительной работы при нахождении двух промежуточных значений одной и той же функции с помощью интерполяционной схемы Эйткена нетрудно объяснить, зная, что эта функция есть $y = \sqrt{x}$. Если на левом конце промежутка $[0, 1.4]$, на котором она задана своими отдельными приближенными значениями, ее производные обращаются в бесконечность, что не позволяет эффективно пользоваться формулой остаточного члена (1.12) для значения $\tilde{x} = 0.1$, близкого к $x_0 = 0$, то при вычислении $y = \sqrt{x}$ в точке 0.5 можно считать, что интерполяция производится на промежутке $[0.4, 1.4]$, где $M_2 = \max(\sqrt{x})'' \approx 1$, $M_3 = \max(\sqrt{x})''' < 4$ и соответственно,

$$|R_1(0.5)| \leq \frac{M_2}{2!} |\pi_2(0.5)| \approx \frac{1}{2} |(0.5 - 0.4)(0.5 - 0.6)| = 0.005,$$

$$|R_2(0.5)| \leq \frac{M_3}{3!} |\pi_3(0.5)| < \frac{4}{6} |(0.5 - 0.4)(0.5 - 0.6)(0.5 - 0.8)| = 0.002.$$

Последняя оценка говорит о том, что квадратичная интерполяция позволила бы вычислить $\sqrt{0.5}$ с точностью практически до третьего знака после запятой, если бы данные в исходной таблице значения функции $y = \sqrt{x}$ имели такую же или бóльшую точность.

Наконец, заметим, что при компьютерных вычислениях заполнение таблиц типа табл.1.2, 1.3, разумеется, не требуется, хотя визуализация этих промежуточных данных много дает для понимания процессов интерполирования. Важно при реализации схемы Эйткена предусмотреть, например, возможность подключения узлов не только последующих, но и предшествующих заданному значению аргумента (из-за конечности таблицы исходных значений функции), причем в средней части таблицы целесообразно чередовать использование информации в последующих и предшествующих узлах.

^{*)} На один полный итерационный шаг можно было сделать меньше, т.е. ограничиться значением $L_5(0.1)$, если учесть, что разница между $P_{0.5}(0.1)$ и $P_{0.6}(0.1)$ составляет величину 0.004, меньшую, чем величина ошибки округления исходных данных (0.005).

§ 1.4. КОНЕЧНЫЕ РАЗНОСТИ

Зададимся целью придать интерполяционной формуле более простой вид, подобный виду широко используемой в математическом анализе формулы Тейлора. Если в интерполяционном многочлене Лагранжа (1.6) все слагаемые однотипны и играют одинаковую роль в образовании результата, хотелось бы иметь такое представление интерполяционного многочлена, в котором, как и в многочлене Тейлора, слагаемые располагались бы в порядке убывания их значимости. Такая структура интерполяционного многочлена позволила бы более просто перестраивать его степень, добавляя или отбрасывая удаленные от начала его записи члены.

Поставленной цели будем добиваться сначала для несколько суженной постановки задачи интерполяции. А именно, будем считать, что интерполируемая функция $y = f(x)$ задана своими значениями $y_0, y_1, \dots, y_n$ на системе **равноотстоящих узлов** $x_0, x_1, \dots, x_n$, т.е. таких, что любой узел x_i этой **сетки** можно представить в виде

$$x_i = x_0 + ih,$$

где $i = 0, 1, \dots, n$, а $h > 0$ — некоторая постоянная величина, называемая **шагом сетки** (таблицы).

Прежде чем строить желаемые интерполяционные формулы, рассмотрим элементы теории **конечных разностей**.

Вычитая из каждого последующего члена конечной последовательности из $n + 1$ чисел $y_0, y_1, \dots, y_n$ предыдущий, образуем n **конечных разностей первого порядка**

$$\Delta y_0 := y_1 - y_0, \quad \Delta y_1 := y_2 - y_1, \quad \dots, \quad \Delta y_{n-1} := y_n - y_{n-1}$$

или, проще, n **первых разностей** данной табличной функции. Из них, в свою очередь, таким же образом можно получить $n - 1$ **конечных разностей второго порядка**, или **вторых разностей**:

$$\begin{aligned} \Delta^2 y_0 &:= \Delta y_1 - \Delta y_0, & \Delta^2 y_1 &:= \Delta y_2 - \Delta y_1, & \dots, \\ \Delta^2 y_{n-2} &:= \Delta y_{n-1} - \Delta y_{n-2}. \end{aligned}$$

Этот процесс построения разностей может быть продолжен, и весь он, очевидно, описывается одной рекуррентной формулой, выражающей **конечную разность k -го порядка** $\Delta^k y_i$ через разности $(k - 1)$ -го порядка:

$$\Delta^k y_i = \Delta^{k-1} y_{i+1} - \Delta^{k-1} y_i, \quad (1.19)$$

где $k = 1, 2, \dots, n$ и $\Delta^0 y_i := y_i$.

В некоторых случаях требуется знать **выражения конечных разностей непосредственно через значения функции**,

лежащей в их основе. Для нескольких первых порядков разностей их можно получить прямой подстановкой:

$$\Delta y_i = y_{i+1} - y_i,$$

$$\Delta^2 y_i = \Delta y_{i+1} - \Delta y_i = y_{i+2} - y_{i+1} - (y_{i+1} - y_i) = y_{i+2} - 2y_{i+1} + y_i,$$

$$\begin{aligned} \Delta^3 y_i &= \Delta^2 y_{i+1} - \Delta^2 y_i = y_{i+3} - 2y_{i+2} + y_{i+1} - (y_{i+2} - 2y_{i+1} + y_i) = \\ &= y_{i+3} - 3y_{i+2} + 3y_{i+1} - y_i, \quad \text{и т.д.} \end{aligned}$$

Подметив закономерность в коэффициентах рассмотренных представлений конечных разностей, записываем общую формулу

$$\Delta^k y_i = \sum_{j=0}^k (-1)^j C_k^j y_{k+i-j}, \quad (1.20)$$

которая может быть строго обоснована методом математической индукции и которая напоминает биномиальное разложение для $(y-1)^k$.

Привлекая определение производной, можно обнаружить определенную связь между конечными разностями и производными. А именно, если учесть, что

$$\lim_{h \rightarrow 0} \frac{y_{i+1} - y_i}{h} = \lim_{h \rightarrow 0} \frac{f(x_i + h) - f(x_i)}{h} = f'(x_i),$$

то можно сказать, что при малых h имеет место приближенное равенство

$$\Delta y_i \approx f'(x_i)h,$$

т.е. первые разности характеризуют первую производную функции $f(x)$, по значениям которой они составлены. Пользуясь этим, имеем для вторых разностей:

$$\begin{aligned} \frac{\Delta^2 y_i}{h^2} &= \frac{\Delta y_{i+1} - \Delta y_i}{h^2} = \frac{\frac{y_{i+2} - y_{i+1}}{h} - \frac{y_{i+1} - y_i}{h}}{h} \approx \\ &\approx \frac{f'(x_i + h) - f'(x_i)}{h} \approx f''(x_i), \end{aligned}$$

т.е. $\Delta^2 y_i \approx f''(x_i)h^2$, и, вообще,

$$\Delta^k y_i \approx f^{(k)}(x_i)h^k. \quad (1.21)$$

Таким образом, на конечные разности можно смотреть как на некоторый аналог производных^{*)}. Отсюда справедливость многих их свойств, одинаковых со свойствами производных.

^{*)} Более подробно связь между производными и конечными разностями изучается далее в главе 6.

Отметим лишь простейшие свойства конечных разностей:

1. конечные разности постоянной равны нулю (очевидно);
2. постоянный множитель у функции можно выносить за знак конечной разности.

Действительно,

$$\Delta(Cy(x)) = Cy(x+h) - Cy(x) = C[y(x+h) - y(x)] = C\Delta y(x)$$

при любых фиксированных x и постоянной C ;

3. конечная разность от суммы двух функций равна сумме их конечных разностей в одной и той же точке.

Свойство проверяется непосредственно: при любых x

$$\begin{aligned}\Delta(u(x) + v(x)) &= u(x+h) + v(x+h) - (u(x) + v(x)) = \\ &= u(x+h) - u(x) + v(x+h) - v(x) = \Delta u(x) + \Delta v(x).\end{aligned}$$

Свойства 2 и 3 характеризуют операцию взятия конечной разности как линейную операцию.

Учитывая роль, какую играют многочлены в теории интерполирования, посмотрим, что представляют собой **конечные разности многочлена**.

Поскольку многочлен в своей канонической форме есть линейная комбинация степенных функций, положим сначала $y = x^n$. Используя биномиальное разложение n -й степени двучлена, получим:

$$\Delta(x^n) = (x+h)^n - x^n = nhx^{n-1} + \frac{n(n-1)}{2!}h^2x^{n-2} + \dots + nh^{n-1}x + h^n,$$

т.е. первая конечная разность степенной функции $y = x^n$ есть многочлен степени $n-1$ со старшим членом nhx^{n-1} . Если взять теперь конечную разность от функции

$$y = a_0x^n + a_1x^{n-1} + \dots + a_{n-1}x + a_n, \quad (1.22)$$

то, в силу линейных свойств Δy , можно записать

$$\Delta y(x) = a_0 \Delta(x^n) + a_1 \Delta(x^{n-1}) + \dots + a_{n-1} \Delta x.$$

Первое слагаемое в этой сумме, как выяснено, есть многочлен $(n-1)$ -й степени, второе, аналогично, — многочлен степени $n-2$, и т.д. Следовательно, первая конечная разность многочлена (1.22) в точке x с шагом h есть тоже многочлен со старшим членом a_0nhx^{n-1} , вторая конечная разность — многочлен со старшим членом $a_0n(n-1)h^2x^{n-2}$, ..., k -я разность — многочлен со старшим членом $a_0n(n-1)\dots(n-k+1)h^kx^{n-k}$.

При $k = n$ получаем постоянную разность n -го порядка

$$\Delta^n y = a_0 n! h^n$$

для многочлена (1.22); конечные разности более высоких порядков, естественно, равны нулю.

Итак, главный вывод из предыдущих рассуждений: *n -е конечные разности многочлена n -й степени постоянны, а $(n + 1)$ -е и все последующие равны нулю.*

Более важным для понимания сути полиномиального интерполирования является утверждение, обратное сделанному выше выводу. А именно, доказано [27], что *если конечные разности n -го порядка некоторой функции $y = y(x)$ постоянны в любой точке x при различных фиксированных шагах h , то эта функция $y(x)$ есть многочлен степени n .*

Для функции $y = f(x)$, заданной таблицей своих значений $y_0, y_1, \dots, y_n$ в узлах $x_0, x_1, \dots, x_n$, где $x_i = x_0 + ih$, конечные разности разных порядков удобно помещать в одну общую таблицу с узлами и значениями функции (последние можно интерпретировать как конечные разности нулевого порядка, см. (1.19)). Эту общую таблицу называют **таблицей конечных разностей**. Заметим, что кроме принятого здесь так называемого диагонального расположения конечных разностей, когда числа в каждом столбце записываются со смещением на полстроки так, как это показано в табл.1.4, часто применяют горизонтальное расположение, где $\Delta y_i, \Delta^2 y_i$ и другие разности с индексом i помещают в одной строке с x_i, y_i .

Таблица 1.4

Диагональная таблица конечных разностей

x_0	y_0	Δy_0	$\Delta^2 y_0$	$\Delta^3 y_0$	$\Delta^4 y_0$	
x_1	y_1	Δy_1	$\Delta^2 y_1$	$\Delta^3 y_1$	$\Delta^4 y_1$	
x_2	y_2	Δy_2	$\Delta^2 y_2$	$\Delta^3 y_2$	$\Delta^4 y_2$	
x_3	y_3	Δy_3	$\Delta^2 y_3$	$\Delta^3 y_3$	$\Delta^4 y_3$	
x_4	y_4	Δy_4	$\Delta^2 y_4$	$\Delta^3 y_4$	$\Delta^4 y_4$	
x_5	y_5	Δy_5	$\Delta^2 y_5$	$\Delta^3 y_5$	$\Delta^4 y_5$	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	

Пример 1.3. Составим таблицу конечных разностей для функции $y = x \ln^2 x$ по ее значениям, вычисленным с тремя знаками после запятой в точках $x_i = 0.4 + 0.2i$, где $i = 0, 1, \dots, 10$. В соответствии с формой, задаваемой табл.1.4, заполняем табл.1.5 всеми возможными для этого случая конечными разностями. Проанализируем ее.

Учитывая связь между конечными разностями и производными соответствующих порядков (см. (1.21)), по смене знаков чисел в столбце Δy_i можно судить о наличии минимума функции в окрестности точки $x = 1$, а положительность всех чисел в столбце $\Delta^2 y_i$ говорит о выпуклости вниз графика данной функции на всем рассматриваемом промежутке $[0.4, 2.4]$.

Далее замечаем, что абсолютные величины конечных разностей сначала убывают с увеличением их порядка, а затем начинают увеличиваться. Это типичное поведение конечных разностей при ограниченной точности задания значений сеточной функции.

Природу наблюдаемого в примере 1.3 поведения модулей конечных разностей нетрудно понять. Если шаг достаточно мал, а данная табличная функция — достаточно гладкая, то сначала происходит естественное убывание $|\Delta^k y_i|$ с увеличением k , в силу упомянутой связи (1.21). Когда эти величины становятся достаточно малыми, большую роль начинают играть продукты взаимодействия исходных ошибок округления (так называемый *шум округлений*).

Что происходит с одной отдельно взятой ошибкой величины ε у значения y_i , можно проследить по табл.1.6. Как видим, с ростом порядка разностей она «расползается» по таблице и увеличивается по абсолютной величине.

Таблица 1.6

Продвижение ошибки по таблице конечных разностей

...	...	...	...	...	...	
x_{i-2}	y_{i-2}		$\Delta^2 y_{i-3}$		$\Delta^4 y_{i-4} + \varepsilon$	
x_{i-1}	y_{i-1}	Δy_{i-2}	$\Delta^2 y_{i-2} + \varepsilon$	$\Delta^3 y_{i-3} + \varepsilon$	$\Delta^4 y_{i-3} - 4\varepsilon$	
x_i	$y_i + \varepsilon$	$\Delta y_{i-1} + \varepsilon$	$\Delta^2 y_{i-1} - 2\varepsilon$	$\Delta^3 y_{i-2} - 3\varepsilon$	$\Delta^4 y_{i-2} + 6\varepsilon$	...
x_{i+1}	y_{i+1}	$\Delta y_i - \varepsilon$	$\Delta^2 y_i + \varepsilon$	$\Delta^3 y_{i-1} + 3\varepsilon$	$\Delta^4 y_{i-1} - 4\varepsilon$	...
x_{i+2}	y_{i+2}	Δy_{i+1}	$\Delta^2 y_{i+1}$	$\Delta^3 y_i - \varepsilon$	$\Delta^4 y_i + \varepsilon$	
...	...	...	...	...	...	

Погрешности, имеющиеся у каждого из данных значений функции, с ростом порядка разностей все больше взаимодействуют.

Из сделанных наблюдений напрашивается следующий вывод. Если какой-то столбец в таблице конечных разностей (в ее эксплуатируемой части) состоит из чисел, абсолютные величины которых составляют всего несколько единиц десятичного знака, являющегося последним в записи исходных значений функции, скажем, не превосходят величины 10ε , где ε — абсолютная погрешность исходных данных, то эти конечные разности и разности всех последующих порядков не несут практически никакой информации о функции, и их не следует использовать. Разности же предшествующего столбца называются **практически постоянными**, и их порядок определяет степень многочлена, которую можно и должно использовать для идеальной в данных условиях полиномиальной интерполяции.

Вспоминая о том, что многочлен k -й степени имеет k -е разности постоянными, а все последующие — нулевыми, приходим к заключению, что *если k -е разности таблицы конечных разностей некоторой функции практически постоянны, то эта функция ведет себя в рассматриваемой области, как многочлен k -й степени*; эту степень и следует применять для интерполирования с наибольшей для данных реалий точностью.

Обратимся к числовой табл.1.5 нашего примера. Видим, что если исключить из рассмотрения верхнюю диагональную строку, то для всей остальной части таблицы третьи разности удовлетворяют условию $|\Delta^3 y_i| \leq 10 \cdot 0.0005$ (где 0.0005 — предельная абсолютная погрешность значений y_i). В такой ситуации разности более высоких порядков не следовало вообще вычислять, а разности второго порядка можно считать практически постоянными, т.е. для подсчета любых промежуточных значений данной функции, за исключением, быть может, тех, которые находятся вблизи узла $x_0 = 0.4$, нужно применять квадратичную интерполяцию.

§ 1.5. КОНЕЧНОРАЗНОСТНЫЕ ИНТЕРПОЛЯЦИОННЫЕ ФОРМУЛЫ

Пусть функция $y = f(x)$ задана на сетке равноотстоящих узлов $x_i = x_0 + ih$, где $i = 0, 1, \dots, n$, и для нее построена таблица конечных разностей 1.4.

В соответствии с тем, что было сказано о направлении модификации интерполяционной формулы Лагранжа в начале предыдущего параграфа, будем строить интерполяционный много-

член $P_n(x)$ в форме

$$P_n(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + \dots \\ \dots + a_n(x - x_0)(x - x_1)\dots(x - x_{n-1}). \quad (1.23)$$

Его $n + 1$ коэффициент $a_0, a_1, \dots, a_n$ будем находить последовательно из $n + 1$ интерполяционных равенств

$$P_n(x_i) = y_i, \quad i = 0, 1, \dots, n.$$

А именно, полагая $i = 0$, т.е. $x = x_0$, в (1.23) имеем $P_n(x_0) = a_0$, а по условию интерполяции $P_n(x_0) = y_0$; следовательно, $a_0 = y_0$.

Далее, при $i = 1$ аналогично получаем равенство

$$a_0 + a_1(x_1 - x_0) = y_1,$$

в которое подставляем уже найденное значение $a_0 = y_0$. Решая это равенство относительно a_1 и используя обозначение конечной разности, получаем

$$a_1 = \frac{y_1 - y_0}{x_1 - x_0} = \frac{\Delta y_0}{h}.$$

Следующий шаг, при $i = 2$, дает:

$$\begin{aligned} a_0 + a_1(x_2 - x_0) + a_2(x_2 - x_0)(x_2 - x_1) &= y_2 \Leftrightarrow \\ y_0 + \frac{\Delta y_0}{h} \cdot 2h + a_2 \cdot 2h \cdot h &= y_2 \Leftrightarrow \\ a_2 = \frac{y_2 - 2y_1 + y_0}{2!h^2} &= \frac{\Delta^2 y_0}{2!h^2} \quad (\text{см. (1.20) при } k = 2). \end{aligned}$$

Полной индукцией можно показать справедливость выражения

$$a_k = \frac{\Delta^k y_0}{k!h^k} \quad \forall k \in \{1, 2, \dots, n\}. \quad (1.24)$$

Подставляя найденные коэффициенты $a_0, a_1, \dots, a_n$ в (1.23), получаем многочлен

$$\begin{aligned} P_n(x) &= y_0 + \frac{\Delta y_0}{h}(x - x_0) + \frac{\Delta^2 y_0}{2!h^2}(x - x_0)(x - x_1) + \dots \\ &\dots + \frac{\Delta^n y_0}{n!h^n}(x - x_0)(x - x_1)\dots(x - x_{n-1}), \end{aligned} \quad (1.25)$$

который называют **первым интерполяционным многочленом Ньютона**.

Учитывая, что каждое слагаемое многочлена (1.25), начиная со второго, содержит множитель $x - x_0$, естественно предположить, что этот многочлен наиболее приспособлен для интерполирования в окрестности узла x_0 (при x , близких к x_0 , $f(x) \approx y_0$). Будем называть узел x_0 **базовым** для многочлена (1.25), и упростим (1.25) введением новой переменной q равенством $q = \frac{x - x_0}{h}$, или (что то же) равенством $x = x_0 + qh$. Так как при любых $i \in \{0, 1, \dots, n\}$

$$x - x_i = x_0 + qh - x_0 - ih = h(q - i),$$

то в результате подстановки этих разностей в (1.25) приходим к **первой интерполяционной формуле Ньютона** в виде

$$f(x) \approx P_n(x_0 + qh) := y_0 + q\Delta y_0 + \frac{q(q-1)}{2!} \Delta^2 y_0 + \dots \\ \dots + \frac{q(q-1)\dots(q-n+1)}{n!} \Delta^n y_0, \quad (1.26)$$

где обозначение $P_n(x_0 + qh)$ указывает не только на n -ю степень многочлена, но и на базовый узел x_0 и связь переменных x и q .

Первая формула Ньютона (1.26) обычно применяется при значениях $|q| < 1$, а именно, для интерполирования вперед, (при $x \in (x_0, x_1)$, т.е. при $q \in (0, 1)$) и экстраполирования назад (при $x < x_0$, т.е. при $q < 0$).

Так как реально степени интерполяционных многочленов бывают не так велики, в то время как таблицы значений функций достаточно обширны, и так как в реальной числовой таблице никаких индексов — номеров узлов нет (см., например, приведенную выше табл.1.5), то за базовый для формулы (1.26) узел x_0 можно принимать узел, ближайший к заданной фиксированной точке x , если за ним имеется достаточное число узлов для построения необходимых для (1.26) разностей. Поскольку в первой формуле Ньютона используются нисходящие диагонали таблицы конечных разностей (см. табл.1.4), то такое смещение узла, принимаемого за базовый, в конце таблицы будет неприемлемо.

Учет этого обстоятельства приводит к потребности в симметричной, в определенном смысле, для (1.26) формулы, которая была бы пригодной для интерполирования в конце таблицы. Для этого, в отличие от (1.23), форма интерполяционного многочлена $P_n(x)$ берется такой, которая предусматривает поочередное подключение узлов в обратном порядке: сначала последний, потом

предпоследний и т.д., т.е.

$$P_n(x) = a_0 + a_1(x - x_n) + a_2(x - x_n)(x - x_{n-1}) + \dots \\ \dots + a_n(x - x_n)(x - x_{n-1}) \dots (x - x_1).$$

Коэффициенты $a_0, a_1, \dots, a_n$ этого многочлена находятся аналогично тому, как они находились для многочлена (1.23), только здесь подстановка узловых точек вместо x и рассмотрение интерполяционных равенств производится тоже в обратном порядке. Полагая $x = x_n, x = x_{n-1}, \dots$, имеем:

$$P_n(x_n) = a_0 = y_n,$$

$$P_n(x_{n-1}) = y_n + a_1(x_{n-1} - x_n) = y_{n-1} \Rightarrow a_1 = \frac{y_n - y_{n-1}}{x_n - x_{n-1}} = \frac{\Delta y_{n-1}}{h},$$

$$P_n(x_{n-2}) = y_n + \frac{\Delta y_{n-1}}{h}(x_{n-2} - x_n) + a_2(x_{n-2} - x_n)(x_{n-2} - x_{n-1}) = y_{n-2}$$

$$\Rightarrow a_2 = \frac{y_{n-2} - y_n + 2\Delta y_{n-1}}{(x_{n-2} - x_n)(x_{n-2} - x_{n-1})} = \frac{y_n - 2y_{n-1} + y_{n-2}}{-2h(-h)} = \frac{\Delta^2 y_{n-2}}{2!h^2}$$

и т.д. В общем случае

$$a_k = \frac{\Delta^k y_{n-k}}{k!h^k} \quad \forall k \in \{1, 2, \dots, n\}.$$

Таким образом получаем **второй интерполяционный многочлен Ньютона**

$$P_n(x) = y_n + \frac{\Delta y_{n-1}}{h}(x - x_n) + \frac{\Delta^2 y_{n-2}}{2!h^2}(x - x_n)(x - x_{n-1}) + \dots \\ \dots + \frac{\Delta^n y_0}{n!h^n}(x - x_n)(x - x_{n-1}) \dots (x - x_1), \quad (1.27)$$

в котором базовым является узел x_n и коэффициенты которого определяются конечными разностями, расположенными на восходящей от y_n диагонали.

Положим в (1.27) $x = x_n + qh$, иначе, введем новую переменную $q = \frac{x - x_n}{h}$ и преобразуем к ней входящие в (1.27) разности:

$$x - x_i = x_n + qh - x_0 - ih = x_0 + nh + qh - x_0 - ih = h(q + n - i).$$

В результате приходим ко **второй интерполяционной формуле Ньютона** вида

$$f(x) \approx P_n(x_n + qh) = y_n + q\Delta y_{n-1} + \frac{q(q+1)}{2!} \Delta^2 y_{n-2} + \dots \\ \dots + \frac{q(q+1)\dots(q+n-1)}{n!} \Delta^n y_0. \quad (1.28)$$

Ее также целесообразно использовать при значениях $|q| < 1$, т.е. в окрестности узла x_n для интерполирования назад (при $q \in (-1, 0)$) и экстраполирования вперед (при $q > 0$).

Наряду с выведенными специально для начала и конца таблицы первой и второй интерполяционными формулами Ньютона, имеется еще несколько формул, рассчитанных на их применение в центральной части таблицы и потому называемых **центральными интерполяционными формулами**. Прежде, чем определять эти формулы, введем понятие центральных разностей.

Будем считать, что узел x_0 расположен в середине таблицы, и нумерация остальных узлов производится, начинаясь с x_0 , с использованием как положительных, так и отрицательных индексов, т.е. считаем $x_i = x_0 + ih$, где $i = 0, \pm 1, \pm 2, \dots$. Тогда центральная часть таблицы конечных разностей будет проиндексирована так, как это показано в табл.1.7. Все подчеркнутые в ней конечные разности (находящиеся с x_0, y_0 в одной строке и на полстроки выше и ниже) называются **центральными разностями**.

Интерполяционный многочлен ищем в форме

$$P(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + \\ + a_3(x - x_{-1})(x - x_0)(x - x_1) + \\ + a_4(x - x_{-1})(x - x_0)(x - x_1)(x - x_2) + \dots, \quad (1.29)$$

предполагающей постепенное подключение узлов x_i : сначала при $i = 0$, затем при $i = 1$, потом при $i = -1$ и т.д., т.е. с двух сторон от x_0 . При этом здесь и далее не будем фиксировать степени многочленов и не будем стремиться выписывать общие и, тем более, последние члены таких многочленов. Как и в предыдущих случаях, коэффициенты a_k ($k = 0, 1, 2, \dots$) находим один за другим последовательной подстановкой в $P(x)$ и в интерполяционные равенства $P(x_i) = y_i$ значений $x = x_0, x_1, x_{-1}, x_2, x_{-2}, \dots$:

$$a_0 = y_0; \quad a_1 = \frac{\Delta y_0}{h}; \quad a_2 = \frac{\Delta^2 y_{-1}}{2!h^2}; \quad a_3 = \frac{\Delta^3 y_{-1}}{3!h^3}; \quad a_4 = \frac{\Delta^4 y_{-2}}{4!h^4}$$

и т.д. Введя новую переменную $q = \frac{x - x_0}{h}$ и выразив через нее

разности $x - x_i = h(q - i)$ для всех $i = 0, \pm 1, \pm 2, \dots$, в результате подстановки этих разностей и выражений коэффициентов в шаблон (1.29), приходим к **первой интерполяционной формуле Гаусса**:

$$\begin{aligned}
 f(x) \approx \bar{P}(x_0 + qh) = & y_0 + q\Delta y_0 + \frac{q(q-1)}{2!} \Delta^2 y_{-1} + \\
 & + \frac{(q+1)q(q-1)}{3!} \Delta^3 y_{-1} + \frac{(q+1)q(q-1)(q-2)}{4!} \Delta^4 y_{-2} + \dots \quad (1.30)
 \end{aligned}$$

Записанные слагаемые легко дополнить следующими, если знать, что в этой формуле используются нижние центральные разности все возрастающих порядков, т.е. те, которые подчеркнуты в табл.1.7 сплошной чертой.

Таблица 1.7

Таблица центральных разностей

x_i	y_i	Δy_i	$\Delta^2 y_i$	$\Delta^3 y_i$	$\Delta^4 y_i$	$\Delta^5 y_i$	$\Delta^6 y_i$	...
...	...	...	...	...	...	...	...	...
x_{-3}	y_{-3}	Δy_{-3}	$\Delta^2 y_{-3}$	$\Delta^3 y_{-3}$	$\Delta^4 y_{-3}$	$\Delta^5 y_{-3}$	$\Delta^6 y_{-3}$	
x_{-2}	y_{-2}	Δy_{-2}	$\Delta^2 y_{-2}$	$\Delta^3 y_{-2}$	$\Delta^4 y_{-2}$	$\Delta^5 y_{-2}$	$\Delta^6 y_{-2}$	
x_{-1}	y_{-1}	Δy_{-1}	$\Delta^2 y_{-1}$	$\Delta^3 y_{-1}$	$\Delta^4 y_{-1}$	$\Delta^5 y_{-1}$	$\Delta^6 y_{-1}$	
x_0	y_0	Δy_0	$\Delta^2 y_0$	$\Delta^3 y_0$	$\Delta^4 y_0$	$\Delta^5 y_0$	$\Delta^6 y_0$	
x_1	y_1	Δy_1	$\Delta^2 y_1$	$\Delta^3 y_1$	$\Delta^4 y_1$	$\Delta^5 y_1$	$\Delta^6 y_1$	
x_2	y_2	Δy_2	$\Delta^2 y_2$	$\Delta^3 y_2$	$\Delta^4 y_2$	$\Delta^5 y_2$	$\Delta^6 y_2$	
x_3	y_3	Δy_3	$\Delta^2 y_3$	$\Delta^3 y_3$	$\Delta^4 y_3$	$\Delta^5 y_3$	$\Delta^6 y_3$	
...	...	...	...	...	...	...	...	
x_{-4}	y_{-4}	Δy_{-4}	$\Delta^2 y_{-4}$	$\Delta^3 y_{-4}$	$\Delta^4 y_{-4}$	$\Delta^5 y_{-4}$	$\Delta^6 y_{-4}$	
x_{-5}	y_{-5}	Δy_{-5}	$\Delta^2 y_{-5}$	$\Delta^3 y_{-5}$	$\Delta^4 y_{-5}$	$\Delta^5 y_{-5}$	$\Delta^6 y_{-5}$	
x_{-6}	y_{-6}	Δy_{-6}	$\Delta^2 y_{-6}$	$\Delta^3 y_{-6}$	$\Delta^4 y_{-6}$	$\Delta^5 y_{-6}$	$\Delta^6 y_{-6}$	

Совершенно аналогично, подключая узлы в другом порядке (после x_0 сначала предшествующий, затем последующий и т.д.,

т.е. $x_0, x_{-1}, x_1, x_{-2}, \dots$), можно построить **вторую интерполяционную формулу Гаусса**

$$f(x) \approx \tilde{P}(x_0 + qh) = y_0 + q\Delta y_{-1} + \frac{(q+1)q}{2!} \Delta^2 y_{-1} + \\ + \frac{(q+1)q(q-1)}{3!} \Delta^3 y_{-2} + \frac{(q+2)(q+1)q(q-1)}{4!} \Delta^4 y_{-2} + \dots, \quad (1.31)$$

использующую верхние центральные разности (подчеркнутые в табл.1.7 пунктирной линией).

Интерполяционные формулы Гаусса служат полуфабрикатами для получения более симметричных, использующих все центральные разности интерполяционных формул.

Так, полусумма первого и второго интерполяционных многочленов Гаусса после преобразований приводит к формуле

$$f(x) \approx P_S(x_0 + qh) = y_0 + q \frac{\Delta y_{-1} + \Delta y_0}{2} + \frac{q^2}{2!} \Delta^2 y_{-1} + \\ + \frac{q(q^2 - 1)}{3!} \cdot \frac{\Delta^3 y_{-2} + \Delta^3 y_{-1}}{2} + \frac{q^2(q^2 - 1)}{4!} \Delta^4 y_{-2} + \dots, \quad (1.32)$$

называемой **интерполяционной формулой Стирлинга^{*)}**.

Если же взять полусумму второго интерполяционного многочлена Гаусса и такого же многочлена, но с нижними индексами, увеличенными на единицу (т.е. с базовой точкой x_1 вместо x_0), то придем к **интерполяционной формуле Бесселя^{**)}**

$$f(x) \approx P_B(x_0 + qh) = \frac{y_0 + y_1}{2} + \left(q - \frac{1}{2}\right) \Delta y_0 + \\ + \frac{q(q-1)}{2!} \cdot \frac{\Delta^2 y_{-1} + \Delta^2 y_0}{2} + \frac{\left(q - \frac{1}{2}\right)q(q-1)}{3!} \Delta^3 y_{-1} + \\ + \frac{q(q-1)(q+1)(q-2)}{4!} \cdot \frac{\Delta^4 y_{-2} + \Delta^4 y_{-1}}{2} + \dots \quad (1.33)$$

В последней формуле обращает на себя внимание тот факт, что она сильно упростится, если в нее подставить значение $q = \frac{1}{2}$, соответствующее значению аргумента $\hat{x} = \frac{1}{2}(x_0 + x_1)$.

^{*)} Стирлинг Джеймс (1692–1770) — шотландский математик.

^{**)} Бессель Фридрих Вильгельм (1784–1846) — немецкий астроном, геодезист и математик.

Этот частный случай формулы Бесселя называют **формулой интерполирования на середину**:

$$f\left(\frac{x_0 + x_1}{2}\right) \approx \frac{y_0 + y_1}{2} - \frac{1}{8} \cdot \frac{\Delta^2 y_{-1} + \Delta^2 y_0}{2} + \\ + \frac{3}{128} \cdot \frac{\Delta^4 y_{-2} + \Delta^4 y_{-1}}{2} - \dots \quad (1.34)$$

Итак, если точка x , в которой нужно найти приближенное значение таблично заданной функции $f(x)$, находится в начале или в конце таблицы, применяется соответственно первая (1.26) или вторая (1.28) формулы Ньютона с таким выбором базовой точки, чтобы значение $|q|$ было как можно меньше. Если точка x находится в середине таблицы, то всегда можно зафиксировать точку x_0 в таблице центральных разностей так,

чтобы $q = \frac{x - x_0}{h}$ либо было по модулю меньше 0.25 и тогда применять интерполяционную формулу Стирлинга (1.32), либо чтобы $q \in [0.25, 0.75]$ и использовать формулу Бесселя (1.33).

Пример 1.4. Пусть требуется для функции $y = f(x)$, заданной в примере 1.3 таблицей нескольких своих значений с тремя знаками после запятой, найти приближенные значения: а) $f(0.5)$; б) $f(1.22)$; в) $f(1.5)$; г) $f(1.94)$; д) $f(2.5)$, записав предварительно соответствующие каждому случаю интерполяционные формулы.

Для решения поставленной задачи учитываем, что значения $f(x)$ заданы в примере 1.3 на сетке равноотстоящих узлов, поэтому здесь можно применить конечноразностную интерполяцию. При этом будем пользоваться уже составленной табл. 1.5 конечных разностей и проведенным ранее ее анализом на выявление оптимальной степени многочлена. Для случаев б)–д) фиксируем вторую степень, для а) — третью. В каждом случае, т.е. для конкретного значения аргумента, выбираем базовый узел, подсчитываем значение вспомогательной переменной q и, в зависимости от положения базового узла и значения q , пользуясь представленными в табл. 1.5 числами, записываем требуемую интерполяционную формулу. Подстановка в нее значения q приводит к искомому значению $f(x)$.

а) При $x = 0.5$ (начало таблицы) полагаем $x_0 = 0.4$; тогда $q = \frac{x - x_0}{h} = \frac{0.5 - 0.4}{0.2} = 0.5$. Соответствующую интерполяционную формулу для аппроксимации $f(x)$ при $x = 0.4 + 0.2q$ с $q \in (-1, 1)$ записываем, глядя на первую интерполяционную формулу Ньютона (1.26) и табл. 1.5:

$$f(x) \approx P_3(0.4 + 0.2q) = 0.336 - 0.179q + \frac{0.062}{2} q(q - 1) + \frac{0.015}{6} q(q - 1)(q - 2).$$

Отсюда получаем искомое значение

$$f(0.5) \approx P_3(0.4 + 0.2 \cdot 0.5) = 0.336 - 0.179 \cdot 0.5 + 0.031 \cdot 0.5 \cdot (-0.5) + 0.0025 \cdot 0.5 \cdot (-0.5) \cdot (-1.5) \approx 0.336 - 0.0895 - 0.0078 + 0.0009 \approx 0.240$$

б) Точка $x = 1.22$ находится в средней части таблицы. Поэтому здесь целесообразно применить формулу Стирлинга или Бесселя. Полагая

$x_0 = 1.2$ и найдя $q = \frac{x - x_0}{h} = \frac{1.22 - 1.2}{0.2} = 0.1$, останавливаемся на формуле Стирлинга (1.32), которая в данном случае имеет вид

$$f(x) \approx P_S(1.2 + 0.2q) = 0.040 + \frac{0.040 + 0.119}{2}q + \frac{0.079}{2}q^2$$

и при $q = 0.1$ приводит к искомому значению

$$f(1.22) \approx 0.040 + 0.0080 + 0.0004 \approx 0.048.$$

в) Здесь, очевидно, напрашивается применение формулы (1.34) интерполирования на середину. Полагая $x_0 = 1.4$, $x_1 = 1.6$, имеем:

$$f(1.5) \approx \frac{0.159 + 0.353}{2} - \frac{1}{8} \cdot \frac{0.075 + 0.075}{2} \approx 0.256 - 0.0094 \approx 0.247.$$

г) Глядя на положение точки $x = 1.94$ в заданной системе узлов табл. 1.5, видим, что для вычисления $f(1.94)$ также возможно применение центральных интерполяционных формул. Положив $x_0 = 1.8$ и вычислив

$q = \frac{x - x_0}{h} = \frac{1.94 - 1.8}{0.2} = 0.7$, на основе (1.33) записываем интерполяционную формулу Бесселя

$$f(x) \approx P_B(1.8 + 0.2q) = \frac{0.622 + 0.961}{2} + 0.339(q - 0.5) + \frac{0.070 + 0.068}{2} \cdot \frac{q(q-1)}{2}.$$

Из нее получаем

$$f(1.94) \approx P_B(1.8 + 0.2 \cdot 0.7) \approx 0.7915 + 0.0678 - 0.0072 \approx 0.852.$$

д) Точка $x = 2.5$ расположена за последним узлом, поэтому для экстраполяции $f(x)$ здесь однозначно следует применить вторую интерполяционную формулу Ньютона (1.28). Считая $x_n = 2.4$ (индекс n здесь используется условно, без придания ему конкретного значения), записываем формулу экстраполяции

$$f(x) \approx P(2.4 + 0.2q) = 1.840 + 0.472q + \frac{0.065}{2}q(q+1),$$

откуда при $q = \frac{x - x_n}{h} = \frac{2.5 - 2.4}{0.2} = 0.5$ находим

$$f(2.5) \approx P(2.4 + 0.2 \cdot 0.5) \approx 1.840 + 0.236 + 0.0244 \approx 2.100.$$

В порядке обсуждения приведенного примера отметим следующее. Во-первых, при записи интерполяционных многочленов не стоит приводить их к канонической форме, ибо тогда их чле-

ны утратят ту информативность, которая в них заложена по построению и которая хорошо видна из рассмотрения промежуточных результатов: налицо убывание роли слагаемых при подсчете каждого из значений а)–д). Во-вторых, вычисление слагаемых интерполяционного результата нет смысла проводить более чем с одним запасным знаком^{*)}, который в конце должен быть отброшен. Об этом говорят приближенность исходных значений и правила приближенных вычислений [2, 18 и др.]. Точность результата интерполирования принципиально не может быть выше, чем точность исходных данных. Вообще, о точности конечно-разностного интерполирования речь пойдет чуть ниже; здесь же заметим, что гладкость данной функции, плотность системы узлов и выбор подходящих параметров интерполяционных формул обеспечили в примере 1.4 вычисление всех требуемых значений, как нетрудно убедиться, с точностью не хуже 0.001, т.е. в грубом смысле все десятичные знаки результатов — верные.

Теперь о том, как могут быть трансформированы остаточный член и его оценки при конечно-разностной интерполяции.

В силу доказанной в § 1.2 единственности интерполяционного многочлена Лагранжа, все построенные здесь конечно-разностные интерполяционные многочлены Ньютона и Гаусса — это всего лишь различные формы его представления^{**)}. Следовательно, для всех этих форм справедливо выражение остаточного члена (1.12), где определенный посредством (1.10) многочлен $\Pi_{n+1}(x)$ для случая равноотстоящих узлов $\{x_i\}_{i=0}^n$ преобразуется

к новой переменной $q = \frac{x - x_0}{h}$ следующим образом:

$$\Pi_{n+1}(x) = \Pi_{n+1}(x_0 + qh) = \prod_{i=0}^n (x_0 + qh - x_0 - ih) = h^{n+1} \prod_{i=0}^n (q - i).$$

Отсюда

$$R_n(x) = R_n(x_0 + qh) = \frac{f^{(n+1)}(\xi)}{(n+1)!} h^{n+1} q(q-1)\dots(q-n). \quad (1.35)$$

Итак, для $f(x) \in C_{[a,b]}^{n+1}$ конечно-разностная интерполяцион-

^{*)} При машинном счете эта рекомендация, разумеется, утрачивает смысл.

^{**) По поводу формул Стирлинга и Бесселя см. далее замечание 1.1.}

ная формула Ньютона или Гаусса степени n с базовым узлом x_0 может быть записана в виде

$$f(x) = P_n(x_0 + qh) + R_n(x_0 + qh), \quad (1.36)$$

где $P_n(x_0 + qh)$ — тот или иной конечноразностный многочлен, построенный по равноотстоящим (с шагом h) узлам $x_0, x_1, \dots, x_n \in [a, b]$, а $R_n(x_0 + qh)$ — остаточный член (1.35), в котором ξ — некоторая неизвестная, но фиксированная (при фиксированном x) точка интервала (a, b) .

Аналогично, при выборе базового узла x_n , т.е. для второй интерполяционной формулы Ньютона, получаем точное представление

$$f(x) = P_n(x_n + qh) + R_n(x_n + qh),$$

где $R_n(x_n + qh) = \frac{f^{(n+1)}(\xi)}{(n+1)!} q(q+1)\dots(q+n)$, а $P_n(x_n + qh)$ — многочлен, определенный в (1.28).

При наличии оценки $|f^{(n+1)}(x)| \leq M_{n+1} \quad \forall x \in [a, b]$ можно уточнить границы абсолютной погрешности конечноразностного интерполирования в конкретной точке и на всем промежутке $[a, b]$ по типу (1.14), (1.15).

Например, для оценки погрешности интерполяции в точке $\tilde{x} = x_0 + \tilde{q}h$ на основании (1.36), (1.35) имеем^{*}

$$|f(\tilde{x}) - P_n(x_0 + \tilde{q}h)| \leq \frac{M_{n+1}}{(n+1)!} h^{n+1} |\tilde{q}(\tilde{q}-1)\dots(\tilde{q}-n)|. \quad (1.37)$$

Если интерполяционная формула $f(x) \approx P_n(x_0 + qh)$ используется для аппроксимации $f(x)$ в точке $\tilde{x}$, расположенной достаточно близко к базовому узлу x_0 справа от него, т.е. если $\tilde{q} \in (0, 1)$, то оценку (1.37) можно существенно упростить.

^{*} В силу отмеченной связи (1.21) между производными и конечными разностями, очень грубо можно заменить M_{n+1} на величину $\frac{\max_i \{|\Delta^{n+1} y_i|\}}{h^{n+1}}$, если, конечно, разности $\Delta^{n+1} y_i$ еще несут какую-то информацию об $f(x)$, т.е. если степень n интерполяционного многочлена берется заниженной по сравнению с той, какой она должна быть для достижения максимальной точности при конкретном порядке практически постоянных разностей.

Это достигается применением неравенства

$$|q(q-1)\dots(q-n)| \leq \frac{n!}{4} \quad \forall q \in (0, 1) \quad \forall n \in \mathbb{N}.$$

(Докажите его методом математической индукции!)

Подстановка последнего неравенства в (1.37) приводит к простой точечной оценке

$$|f(\tilde{x}) - P_n(x_0 + \tilde{q}h)| \leq \frac{M_{n+1}}{4(n+1)} h^{n+1} \quad (\text{при } \tilde{q} \in (0, 1)), \quad (1.38)$$

подчеркивающей степенную зависимость точности интерполирования от малости шага таблицы. Если шаг h мал и с ростом n не происходит слишком быстрого роста $|f^{(n+1)}(x)|$ при $x \in (a, b)$, то, как видно из (1.38), повышение степени n интерполяционного многочлена $P_n(x_0 + qh)$ за счет множителя h^{n+1} влечет уменьшение погрешности интерполяции (здесь имеется в виду только методическая погрешность, т.е. не учитывается точность задания значений функции и шум округлений). В этом случае можно говорить о точечной сходимости процесса конечноразностного интерполирования.

С равномерной сходимостью дело обстоит сложнее. Существуют примеры бесконечно дифференцируемых функций, для которых максимальная погрешность при интерполировании на заданном отрезке по системе равноотстоящих узлов не стремится к нулю при $n \rightarrow \infty$ ($h \rightarrow 0$). Таким примером служит функция

$y = \frac{1}{1 + 25x^2}$, впервые рассмотренная с этой целью Рунге^{*)}. Установлено [2, 71, 74 и др.], что при ее интерполировании на отрезке $[-1, 1]$ по системе равноотстоящих узлов с шагом $h = \frac{2}{n}$ многочленами $P_n(x)$ имеет место

$$\lim_{n \rightarrow \infty} \max_{x \in [-1, 1]} |y - P_n(x)| = \infty.$$

Завершая разговор о конечноразностной интерполяции, отметим, что наряду с уже построенными здесь интерполяционными многочленами, можно строить ряд других. Для этого существует некая схема, называемая **диаграммой Фрезера** [11 и др.], включающая в себя известные интерполяционные многочлены и

^{*)} Рунге Карл Давид Тольме (1856–1927) — немецкий физик и математик.

служащая источником получения новых, предполагающих использование зигзагообразных путей продвижения по таблице конечных разностей.

Замечание 1.1. Формулы Стирлинга (1.32) и Бесселя (1.33) лишь условно можно назвать интерполяционными. Это связано со способом их получения — суммированием с коэффициентами 0.5 действительно интерполяционных формул Гаусса. Легко видеть, что, например, в линейном случае получаемая в результате такого суммирования многочленов $\bar{P}_1(x_0 + qh) = y_0 + q\Delta y_0$ и $\tilde{P}_1(x_0 + qh) = y_0 + q\Delta y_{-1}$ формула Стирлинга

$$f(x) \approx y_0 + q \frac{\Delta y_{-1} + \Delta y_0}{2} = y_0 + \frac{y_1 - y_{-1}}{2h} (x - x_0) =: P_s^1$$

определяется значениями данной функции $y = f(x)$ в трех точках x_{-1} , x_0 и x_1 (отсюда априори ее более высокая точность), но условие интерполяции выполняется лишь в одной из них, x_0 (рис. 1.2). Забегая вперед, ска-

жем, что $\frac{y_1 - y_{-1}}{2h}$ лучше аппроксимирует $f'(x)$, чем $\frac{y_1 - y_0}{h}$ или

$\frac{y_0 - y_{-1}}{h}$ (ср. (6.18) с (6.14) и (6.15)), т.е. формула Стирлинга ближе к формуле Тейлора, чем формула Гаусса или Ньютона, а значит многочлен Стирлинга лучше аппроксимирует функцию $f(x)$ в окрестности точки x_0 .

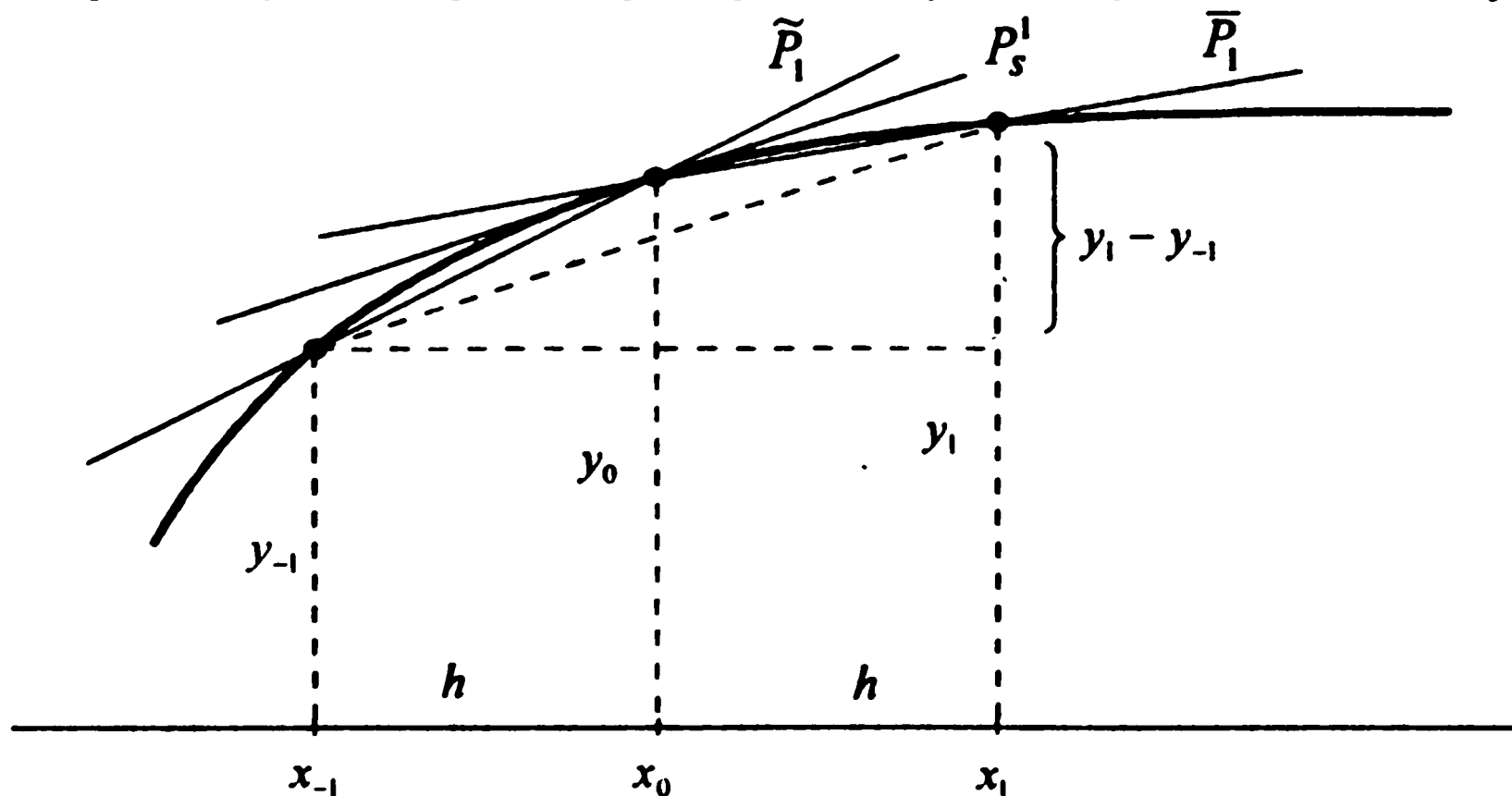


Рис. 1.2. Геометрическая интерпретация многочлена Стирлинга первой степени.

Формула Бесселя нулевого порядка

$$f(x) \approx \frac{y_0 + y_1}{2}$$

определяется значениями функции $y = f(x)$ в узлах x_0 и x_1 и не является точной ни в одном из них; следовательно, многочлен Бесселя нулевой сте-

пени $P_B^0 := \frac{y_0 + y_1}{2}$ не удовлетворяет условиям интерполяции (1.3) и называть его интерполяционным можно лишь в некотором более широком смысле.

В связи с отмеченной особенностью формул Стирлинга и Бесселя, выделяющей их из множества других, эквивалентных интерполяционной формуле Лагранжа конечноразностных интерполяционных формул, для них должны применяться отличные от (1.35) формулы остаточных членов; их можно найти, например, в [11].

§ 1.6. ИНТЕРПОЛЯЦИОННАЯ ФОРМУЛА НЬЮТОНА ДЛЯ НЕРАВНООТСТОЯЩИХ УЗЛОВ

Для построения интерполяционных формул, имеющих перед классической интерполяционной формулой Лагранжа (1.6) преимущества, какими обладают конечноразностные формулы, и применимых в более общем по сравнению с последними случае произвольного расположения упорядоченных несовпадающих узлов $x_0, x_1, \dots, x_n$ на промежутке $[a, b]$, вместо конечных разностей используют разделенные разности, или иначе, разностные отношения.

Через значения функции $f(x_0), f(x_1), \dots, f(x_n)$ сначала определяют *разделенные разности первого порядка*:

$$\begin{aligned} f(x_0; x_1) &:= \frac{f(x_1) - f(x_0)}{x_1 - x_0}, \\ f(x_1; x_2) &:= \frac{f(x_2) - f(x_1)}{x_2 - x_1}, \\ &\dots \dots \dots \\ f(x_{n-1}; x_n) &:= \frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}}. \end{aligned}$$

На этих разностях базируются *разделенные разности второго порядка*:

$$\begin{aligned} f(x_0; x_1; x_2) &:= \frac{f(x_1; x_2) - f(x_0; x_1)}{x_2 - x_0}, \\ &\dots \dots \dots \\ f(x_{n-2}; x_{n-1}; x_n) &:= \frac{f(x_{n-1}; x_n) - f(x_{n-2}; x_{n-1})}{x_n - x_{n-2}}, \end{aligned}$$

и т.д. Таким образом, если определены k -е разностные отношения $f(x_i; x_{i+1}; \dots; x_{i+k})$, то $(k+1)$ -е определяются через них равенством

$$f(x_{i-1}; x_i; \dots; x_{i+k}) = \frac{f(x_i; x_{i+1}; \dots; x_{i+k}) - f(x_{i-1}; x_i; \dots; x_{i+k-1})}{x_{i+k} - x_{i-1}}. \quad (1.39)$$

Легко проверить, что операция взятия разделенной разности, как и в случае конечной разности, аддитивна и однородна, т.е. линейна. Кроме того, разделенная разность есть симметрическая функция своих аргументов, что позволяет у разделенной разности менять аргументы местами.

Доказательство симметричности опирается на представление k -й разделенной разности через значения функции в узлах, имеющее вид

$$f(x_i; x_{i+1}; \dots; x_{i+k}) = \sum_{j=0}^k \frac{f(x_{i+j})}{(x_{i+j} - x_i) \dots (x_{i+j} - x_{i+j-1})(x_{i+j} - x_{i+j+1}) \dots (x_{i+j} - x_{i+k})}. \quad (1.40)$$

При $k=1$ справедливость выражения (1.40) очевидна:

$$f(x_i; x_{i+1}) = \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i} = \frac{f(x_{i+1})}{x_{i+1} - x_i} + \frac{f(x_i)}{x_i - x_{i+1}}.$$

Для произвольных натуральных k равенство (1.40) доказывается на основе (1.39) по индукции [11].

Важно отметить, что как и для конечных разностей, взятие разделенной разности многочлена понижает на единицу его степень. Это следует из рассмотрения разделенной разности $f(x_i; x_{i+1})$ степенной функции $f(x) = x^n$:

$$f(x_i; x_{i+1}) = \frac{x_{i+1}^n - x_i^n}{x_{i+1} - x_i} = x_{i+1}^{n-1} + x_{i+1}^{n-2} \cdot x_i + \dots + x_{i+1} \cdot x_i^{n-2} + x_i^{n-1}.$$

Отсюда приходим к заключению, что разделенные разности n -го порядка многочлена n -й степени

$$P_n(x) = a_0 x^n + a_1 x^{n-1} + \dots + a_{n-1} x + a_n$$

постоянны^{*)}.

Следовательно, как и в конечноразностном случае, анализируя таблицу разделенных разностей (табл.1.8), по порядку

^{*)} Показано [11], что $P_n(x; x+h_1; \dots; x+h_n) \equiv a_0$, а $(n+1)$ -е и все последующие разделенные разности равны нулю.

Таблица 1.8

Таблица разделенных разностей

x_i	$f(x_i)$	$f(x_i; x_{i+1})$	$f(x_i; x_{i+1}; x_{i+2})$	$f(x_i; x_{i+1}; x_{i+2}; x_{i+3})$	$f(x_i; x_{i+1}; x_{i+2}; x_{i+3}; x_{i+4})$	...
x_0	$f(x_0)$					
x_1	$f(x_1)$	$f(x_0; x_1)$	$f(x_0; x_1; x_2)$			
x_2	$f(x_2)$	$f(x_1; x_2)$	$f(x_1; x_2; x_3)$	$f(x_0; x_1; x_2; x_3)$	$f(x_0; x_1; x_2; x_3; x_4)$	
x_3	$f(x_3)$	$f(x_2; x_3)$	$f(x_2; x_3; x_4)$	$f(x_1; x_2; x_3; x_4)$	$f(x_1; x_2; x_3; x_4; x_5)$	...
x_4	$f(x_4)$	$f(x_3; x_4)$	$f(x_3; x_4; x_5)$	$f(x_2; x_3; x_4; x_5)$	$f(x_2; x_3; x_4; x_5; x_6)$	...
x_5	$f(x_5)$	$f(x_4; x_5)$	$f(x_4; x_5; x_6)$	$f(x_3; x_4; x_5; x_6)$	...	
x_6	$f(x_6)$	$f(x_5; x_6)$	...	...		
...	...					

почти совпадающих разделенных разностей можно сделать вывод о предпочтительной степени многочлена, подходящего для интерполирования данной функции.

В основу построения интерполяционного многочлена по типу первого интерполяционного многочлена Ньютона (1.25), но для случая неравных промежутков между узлами x_i , положим следующие рассуждения.

Пусть $\varphi(x)$ — некоторая функция с известными значениями в узлах $x_0, x_1, \dots$, а x — произвольная фиксированная точка. По определению разделенной разности первого порядка имеем

$$\varphi(x; x_0) = \frac{\varphi(x_0) - \varphi(x)}{x_0 - x} = \frac{\varphi(x) - \varphi(x_0)}{x - x_0},$$

откуда

$$\varphi(x) = \varphi(x_0) + \varphi(x; x_0)(x - x_0). \quad (1.41)$$

Для разделенной разности второго порядка по точкам x, x_0, x_1 записываем представление

$$\varphi(x; x_0; x_1) = \frac{\varphi(x_0; x_1) - \varphi(x; x_0)}{x_1 - x} = \frac{\varphi(x; x_0) - \varphi(x_0; x_1)}{x - x_1},$$

следствием которого является выражение

$$\varphi(x; x_0) = \varphi(x_0; x_1) + \varphi(x; x_0; x_1)(x - x_1).$$

Подставляя его в (1.41), приходим к равенству

$$\varphi(x) = \varphi(x_0) + \varphi(x_0; x_1)(x - x_0) + \varphi(x; x_0; x_1)(x - x_0)(x - x_1).$$

Формально, на основе определяющего разделенные разности рекуррентного соотношения (1.39), этот процесс может быть продолжен. В результате можно записать формулу, описывающую своеобразное разложение $\varphi(x)$ по произведениям разностей $(x - x_i)$, коэффициентами в котором являются разделенные разности различных порядков:

$$\begin{aligned} \varphi(x) = & \varphi(x_0) + \varphi(x_0; x_1)(x - x_0) + \varphi(x_0; x_1; x_2)(x - x_0)(x - x_1) + \dots \\ & \dots + \varphi(x_0; x_1; \dots; x_n)(x - x_0) \dots (x - x_{n-1}) + \\ & + \varphi(x; x_0; x_1; \dots; x_n)(x - x_0) \dots (x - x_{n-1})(x - x_n). \end{aligned} \quad (1.42)$$

Если $\varphi(x) \equiv P_n(x)$ — многочлен степени n , то процесс подобного разложения исчерпывается. Разложение будет состоять из $n + 1$ слагаемого, и все они будут иметь конкретные коэффициенты, так как последняя, содержащая x , разделенная разность

в (1.42), т.е. $\varphi(x; x_0; \dots; x_n) \equiv P_n(x; x_0; \dots; x_n)$ имеет $(n+1)$ -й порядок и, значит, равна нулю. Таким образом, для любого многочлена степени n справедливо тождество

$$\begin{aligned} P_n(x) = & P_n(x_0) + P_n(x_0; x_1)(x - x_0) + \\ & + P_n(x_0; x_1; x_2)(x - x_0)(x - x_1) + \dots \\ & \dots + P_n(x_0; x_1; \dots; x_n)(x - x_0)(x - x_1) \dots (x - x_{n-1}). \end{aligned}$$

Предположим, что этот многочлен $P_n(x)$ является интерполяционным для некоторой функции $f(x)$. Тогда во всех узлах $x_0, x_1, \dots, x_n$ он должен иметь одинаковые с ней значения, а значит должны быть одинаковыми и их разделенные разности. Отсюда приходим к **интерполяционной формуле Ньютона для неравноотстоящих узлов**:

$$\begin{aligned} f(x) \approx P_n(x) := & f(x_0) + f(x_0; x_1)(x - x_0) + \\ & + f(x_0; x_1; x_2)(x - x_0)(x - x_1) + \dots \\ & \dots + f(x_0; x_1; \dots; x_n)(x - x_0)(x - x_1) \dots (x - x_{n-1}). \end{aligned} \quad (1.43)$$

Подставив $f(x)$ вместо $\varphi(x)$ в (1.42), с учетом (1.43) получаем точное равенство

$$f(x) = P_n(x) + f(x; x_0; \dots; x_n)(x - x_0) \dots (x - x_{n-1})(x - x_n),$$

второе слагаемое в котором может рассматриваться в качестве остаточного члена, т.е.

$$R_n(x) := f(x) - P_n(x) = f(x; x_0; \dots; x_n)\Pi_{n+1}(x), \quad (1.44)$$

где $\Pi_{n+1}(x)$ — многочлен, введенный ранее посредством (1.10).

Поскольку для вычисления разности $f(x; x_0; \dots; x_n)$ требуется знание значения $f(x)$ наряду с известными значениями $f(x_0), \dots, f(x_n)$, представляемое формулой (1.44) выражение $R_n(x)$ фактически можно использовать только для оценивания погрешности интерполирования по формуле (1.43) через максимальные величины модулей разделенных разностей $(n+1)$ -го порядка (если в них содержится достаточно незашумленной информации) или для получения других выражений остаточного члена при тех или иных предположениях о данной функции. В частности, если функция $f(x)$ имеет $(n+1)$ -ю производную, то остаточный член (1.44) может быть преобразован к виду (1.12), согласно утверждению о единственности интерполяционного многочлена Лагранжа, одной из форм представления которого является (1.43).

Таблица 1.9

Разделенные разности функции $f(x)=\cos(e^{-x})$

x_i	$f(x_i)$	$f(x_i;x_{i+1})$	$f(x_i;\dots;x_{i+2})$	$f(x_i;\dots;x_{i+3})$	$f(x_i;\dots;x_{i+4})$	$f(x_i;\dots;x_{i+5})$	$f(x_i;\dots;x_{i+6})$	$f(x_i;\dots;x_{i+7})$
0.0	0.5403							
		0.7140						
0.2	0.6831		-0.5046					
		0.4617		0.2124				
0.5	0.8216		-0.3134		-0.0576			
		0.2423		0.1317		0.0097		
0.9	0.9185		-0.1554		-0.0382		-0.0006	
		0.1024		0.0629		0.0080		-0.0002
1.4	0.9697		-0.0610		-0.0182		-0.0012	
		0.0353		0.0229		0.0036		
2.0	0.9909		-0.0197		-0.0055			
		0.0097		0.0060				
2.7	0.9977		-0.0041					
		0.0016						
4.0	0.9998							

При практическом использовании интерполяционной формулы (1.43) приходится полагаться на убывание модулей слагаемых в $P_n(x)$ при увеличении номера слагаемого. Такое убывание обычно происходит до некоторых пор; затем начинается рост их модулей из-за влияния ошибок округления.

Пример 1.5. Проследим процесс формирования результата интерполирования функции $f(x) = \cos(e^{-x})$ в точке $x = 0.3$ с помощью нескольких ее приближенных значений, заданных на неравномерной сетке. Эти заданные значения и вычисленные по ним разделенные разности всех возможных порядков представлены в табл.1.9.

Полагая в (1.43) $x = 0.3$, $x_0 = 0.2$ и, далее, выбирая из табл.1.9 нужные для применения (1.43) данные, получаем:

$$\begin{aligned} f(0.3) &\approx 0.6831 + 0.4617 \cdot 0.1 - 0.3134 \cdot 0.1 \cdot (-0.2) + 0.1317 \cdot 0.1 \cdot (-0.2) \cdot (-0.6) - \\ &- 0.0382 \cdot 0.1 \cdot (-0.2) \cdot (-0.6) \cdot (-1.1) + 0.0080 \cdot 0.1 \cdot (-0.2) \cdot (-0.6) \cdot (-1.1) \cdot (-1.7) \approx \\ &\approx 0.6831 + 0.04617 + 0.00627 + 0.00158 + 0.00050 + 0.00018 \approx 0.7378 \end{aligned}$$

Найденное значение отличается от истинного значения $\cos(e^{-0.3})$ не более, чем на 0.0001, что в условиях точности исходных данных в пределах 0.00005 можно считать вполне хорошим результатом. Обратим внимание на предсказанное постепенное убывание роли последующих слагаемых в его формировании.

Заметим, что при использовании разделенных разностей, как и при конечноразностной интерполяции, можно менять последовательность подключения узлов и получать интерполяционные формулы для неравных промежутков, отличные от (1.43), но имеющие такую же структуру.

§ 1.7. ОБРАТНОЕ ИНТЕРПОЛИРОВАНИЕ

Под *задачей обратного интерполирования* понимается задача нахождения приближенного значения $\hat{x}$ аргумента x таблично заданной функции $f(x)$, соответствующего некоторому известному значению $\hat{y}$ этой функции. При этом, естественно, предполагается, что $\hat{y}$ не совпадает ни с одним из данных значений $y_i = f(x_i)$ (и значит, $\hat{x}$ не совпадает ни с одним из узлов x_i) и, в то же время, достаточно хорошо «вписывается» в таблицу значений y_i .

Для того, чтобы говорить о теоретически однозначной разрешимости задачи обратного интерполирования, нужно потребовать выполнения основного условия существования обратной функции $x = f^{-1}(y)$ — монотонность данной сеточной функции.

Условимся далее считать, что такая монотонность имеет место либо во всей исходной таблице значений y_i , либо в некоторой ее части (используемой для решения обратной задачи).

Так как функция $y = f(x)$, вообще говоря, не может быть восстановлена точно по нескольким своим значениям (за определенными исключениями), то и обратная задача заведомо может решаться лишь приближенно.

Формально простейший из приемов решения задачи обратной интерполяции заключается в перемене ролями функции и аргумента и применения интерполяционной формулы Лагранжа, т.е. непосредственное вычисление $\hat{x}$ по формуле (ср. (1.6))

$$\hat{x} \approx \sum_{i=0}^n \frac{(\hat{y} - y_0)(\hat{y} - y_1) \dots (\hat{y} - y_{i-1})(\hat{y} - y_{i+1}) \dots (\hat{y} - y_n)}{(y_i - y_0)(y_i - y_1) \dots (y_i - y_{i-1})(y_i - y_{i+1}) \dots (y_i - y_n)} \cdot x_i. \quad (1.45)$$

Такой подход обладает характерными для применений формулы Лагранжа недостатками (обсужденными ранее, см., в частности, § 1.3) и привлекателен лишь в случаях, когда степень n интерполяционного многочлена невысока и заранее известна.

Например, им можно воспользоваться для следующего вывода *метода обратной линейной интерполяции* (иначе, метода хорд) решения скалярных уравнений вида $f(x) = 0$ [18 и др.].

Пусть для непрерывной функции $y = f(x)$ известны две точки x_0, x_1 такие, что $f(x_0) \cdot f(x_1) < 0$, т.е. между ними имеется точка $\hat{x}$ такая, что $\hat{y} := f(\hat{x}) = 0$. Приближенно эту точку можно найти обратной линейной интерполяцией, подставив в (1.45) $n = 1$ и $\hat{y} = 0$:

$$\hat{x} \approx \frac{0 - y_1}{y_0 - y_1} x_0 + \frac{0 - y_0}{y_1 - y_0} x_1.$$

Обозначая правую часть этого приближенного равенства через x_2 и учитывая, что $y_0 = f(x_0)$, $y_1 = f(x_1)$, полученному придаем вид

$$\hat{x} \approx x_2 := \frac{x_0 f(x_1) - x_1 f(x_0)}{f(x_1) - f(x_0)}.$$

На основе последней формулы (возможно, более узнаваемой в виде

$x_2 = x_1 - \frac{f(x_1)(x_1 - x_0)}{f(x_1) - f(x_0)}$) строится итерационный процесс получения по-

следовательных приближений x_k к корню $\hat{x}$ уравнения $f(x) = 0$ (см. рис. 1.3, а также [18]).

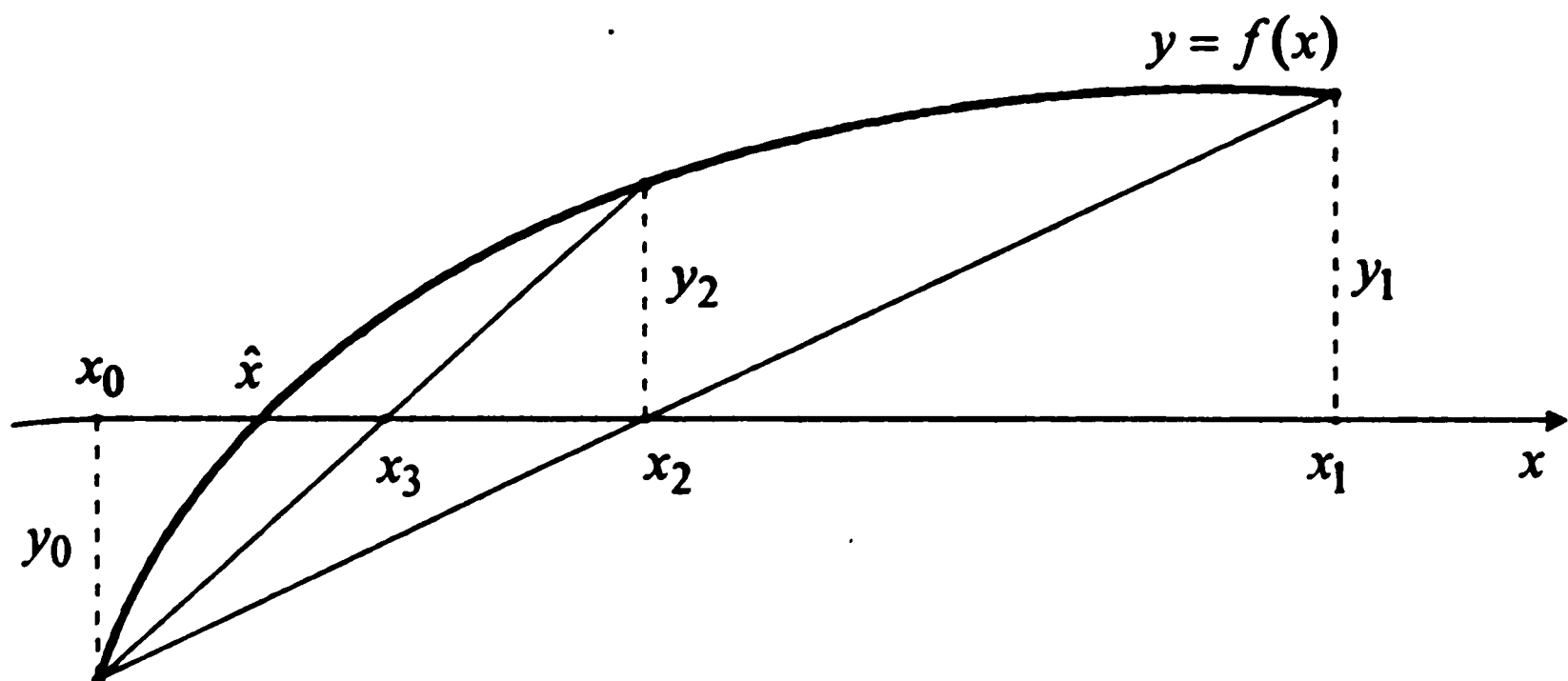


Рис. 1.3. Приближения к корню $\hat{x}$ уравнения $f(x)=0$ методом обратной линейной интерполяции

Используя сразу большее, чем в предыдущем случае, количество точечной информации о данной функции $f(x)$, т.е. подменяя функцию $f(x)$ интерполяционным многочленом $P_n(x)$ достаточно высокой степени, можно за один шаг найти удовлетворительное приближение к искомому значению $\hat{x} = f^{-1}(\hat{y})$ и, в частности, при $\hat{y} = 0$ с хорошей точностью найти корень уравнения $f(x) = 0$.

Если сопрягать точность получения искомого значения $\hat{x} = f^{-1}(\hat{y})$ с построением соответствующего интерполяционного многочлена подходящей (неизвестной заранее) степени, то для этой цели более пригодны интерполяционные формулы, в которых можно постепенно наращивать точность за счет последовательного подключения все новых узлов. Рассмотрим, как это может быть реализовано в случае, когда информация об $f(x)$ берется на системе равноотстоящих узлов.

Предположим, что число $\hat{y}$ близко к $y_0 = f(x_0)$, например, вписывается между y_0 и $y_1 = f(x_1)$, и пусть известно достаточное количество значений y_i функции $f(x)$ в точках $x_i = x_0 + ih$ при $i = 0, 1, 2, \dots$. Тогда за основу может быть принята первая интерполяционная формула Ньютона (1.26) с базовым узлом x_0 . В силу выдвинутого ранее требования о монотонности $f(x)$, искомая точка $\hat{x}$ должна быть близка к x_0 ; следовательно, приближенное значение $\hat{x}$ будет найдено, если удастся найти такое (вообще говоря, малое, укладывающееся в границы интервала $(0, 1)$)

значение $\hat{q} = \frac{\hat{x} - x_0}{h}$, которое является приближенным корнем

уравнения

$$y_0 + q\Delta y_0 + \frac{q(q-1)}{2!}\Delta^2 y_0 + \dots + \frac{q(q-1)\dots(q-n+1)}{n!}\Delta^n y_0 = \hat{y}. \quad (1.46)$$

Учитывая специфику уравнения (1.46), придадим ему вид задачи о неподвижной точке

$$q = \frac{\hat{y} - y_0}{\Delta y_0} - \frac{\Delta^2 y_0}{2!\Delta y_0} q(q-1) - \dots - \frac{\Delta^n y_0}{n!\Delta y_0} q(q-1)\dots(q-n+1).$$

К последнему можно применить метод простых итераций [18], т.е. находить последовательные приближения q_k к $\hat{q}$ по формуле

$$q_{k+1} = \frac{\hat{y} - y_0}{\Delta y_0} - \frac{\Delta^2 y_0}{2!\Delta y_0} q_k(q_k - 1) - \dots - \frac{\Delta^n y_0}{n!\Delta y_0} q_k(q_k - 1)\dots(q_k - n + 1) \quad (1.47)$$

при $k = 0, 1, 2, \dots$, начиная с $q_0 = \frac{\hat{y} - y_0}{\Delta y_0}$. Степень n интерполя-

ционного многочлена здесь фиксируется в соответствии с поведением конечных разностей (не обязательно сразу максимальная; можно ее наращивать постепенно от итерации к итерации), а число итераций k , при котором следует положить $\hat{q} \approx q_k$ определяется практическим совпадением q_k с q_{k-1} в пределах тех знаков, на которые можно рассчитывать при той или иной точности и разреженности исходной дискретной информации об $f(x)$. После того, как при некотором $k \geq 1$ будет принято $\hat{q} \approx q_k$, считаем $\hat{x} \approx x_0 + \hat{q}h$.

Для решения задачи обратной интерполяции можно составить и другие аналогичные (1.46) уравнения с помощью подходящих для того или иного случая разностных интерполяционных формул. Например, на базе интерполяционной формулы Ньютона для неравноотстоящих промежутков (1.43) можно построить итерационный процесс вычисления последовательных приближений $x^{(k)}$ к искомому значению $\hat{x}$ по формуле

$$x^{(k+1)} = \varphi(x^{(k)}); \quad k = 0, 1, 2, \dots; \quad x^{(0)} = x_0,$$

где

$$\varphi(x) := x_0 - \frac{1}{f(x_0; x_1)} [f(x_0) - \hat{y} + f(x_0; x_1; x_2)(x - x_0)(x - x_1) + \dots + f(x_0; x_1; \dots; x_n)(x - x_0)(x - x_1)\dots(x - x_{n-1})].$$

Процесс подключения слагаемых в правой части, содержащих разделенные разности все более высоких порядков, здесь также может быть организован последовательным.

Пример 1.6. Пусть известно несколько приближенных значений функции $f(x) = \frac{\ln(x \ln(1+x))}{x}$, и по ним построена таблица конечных разностей (табл.1.10). Требуется найти приближенно корень уравнения $f(x) = 0$.

Полагая $\hat{y} = 0$, $x_0 = 1.2$ и $n = 3$, в соответствии с формулой последовательных приближений (1.47) и данными табл.1.10, находим:

$$q_0 = \frac{0 - (-0.046)}{0.107} \approx 0.430;$$

$$q_1 = 0.430 - \frac{(-0.023)}{2 \cdot 0.107} \cdot 0.430 \cdot (0.430 - 1) - \frac{0.006}{6 \cdot 0.107} \cdot 0.430 \cdot (0.430 - 1)(0.430 - 2) \approx \\ \approx 0.430 - 0.026 - 0.004 = 0.400;$$

$$q_2 = 0.430 - \frac{(-0.023)}{2 \cdot 0.107} \cdot 0.400 \cdot (0.400 - 1) - \frac{0.006}{6 \cdot 0.107} \cdot 0.400 \cdot (0.400 - 1)(0.400 - 2) \approx \\ \approx 0.430 - 0.026 - 0.004 = 0.400.$$

Совпадение q_2 с q_1 в третьем знаке после запятой говорит о том, что можно принять $\hat{q} \approx q_2 \approx 0.400$. Учитывая, что $h = 0.1$, в силу связи $\hat{x} = x_0 + qh$, корень уравнения $\frac{\ln(x \ln(1+x))}{x} = 0$ считаем равным

$$\hat{x} \approx 1.2 + 0.400 \cdot 0.1 = 1.240.$$

Таблица 1.10

Таблица конечных разностей функции $y = \frac{\ln(x \ln(1+x))}{x}$

x_i	y_i	Δy_i	$\Delta^2 y_i$	$\Delta^3 y_i$	$\Delta^4 y_i$
1.2	-0.046	0.107			
1.3	0.061				
1.4	0.145	0.084	-0.023	0.006	-0.003
		0.067	-0.017		
1.5	0.212	0.053	-0.014		
1.6	0.265				

§ 1.8. ИНТЕРПОЛЯЦИЯ С КРАТНЫМИ УЗЛАМИ

Рассмотрим задачу полиномиальной интерполяции функции $y = f(x)$ в более общей постановке.

Пусть на промежутке $[a, b] \subseteq D(f)$ расположены $m+1$ несовпадающих узлов $x_0, x_1, \dots, x_m$, и пусть в этих точках известны значения $y_0 = f(x_0), y_1 = f(x_1), \dots, y_m = f(x_m)$ данной функции, а также некоторые ее производные (максимальный порядок производных в разных узлах различен; в каких-то узлах производные могут быть вовсе неизвестны). Такие узлы будем называть **кратными узлами**. Конкретнее, будем считать, что заданы:

$$\begin{aligned} &\text{в узле } x_0 \text{ значения } y_0, y'_0, \dots, y_0^{(k_0-1)}, \\ &\text{в узле } x_1 \text{ значения } y_1, y'_1, \dots, y_1^{(k_1-1)}, \\ &\dots\dots\dots \\ &\text{в узле } x_m \text{ значения } y_m, y'_m, \dots, y_m^{(k_m-1)}; \end{aligned} \tag{1.48}$$

тогда **кратность** узла x_0 считается равной k_0 , узла x_1 — k_1 , ..., узла x_m — k_m .

Предполагая, что **суммарная кратность узлов** есть

$$k_0 + k_1 + \dots + k_m = n + 1, \tag{1.49}$$

ставим задачу построения многочлена $H_n(x)$ степени n (не выше n) такого, что

$$H_n^{(j)}(x_i) = y_i^{(j)} \quad \forall i \in \{0, 1, \dots, m\} \quad \forall j \in \{0, 1, \dots, k_i - 1\}, \tag{1.50}$$

где $m \geq 0$, $y_i^{(j)} := f^{(j)}(x_i)$ — заданные посредством (1.48) значения функции $f(x)$ и ее производных и по определению считается $H_n^{(0)}(x_i) := H_n(x_i)$, $y_i^{(0)} := y_i$. Многочлен $H_n(x)$ будем называть **интерполяционным многочленом Эрмита^{*}**, а совокупность требований (1.50) — **условиями эрмитовой интерполяции**.

Формально можно считать, что нахождение такого многочлена состоит в том, чтобы однозначно определить $n+1$ коэф-

^{*} Эрмит Шарль (1822–1901) — французский математик. Принятое обозначение многочлена $H_n(x)$ связано с французским написанием фамилии Hermite.

коэффициент $a_0, a_1, \dots, a_n$ его канонического представления

$$H_n(x) = a_0 x^n + a_1 x^{n-1} + \dots + a_{n-1} x + a_n \quad (1.51)$$

из условий (1.50). В силу предположения (1.49) о суммарной кратности узлов эрмитовой интерполяции, совокупность требований (1.50) можно рассматривать как систему из $n+1$ уравнения относительно $n+1$ неизвестного — коэффициентов a_k многочлена (1.51):

$$\begin{array}{llll} H_n(x_0) = y_0; & H'_n(x_0) = y'_0; & \dots & H_n^{(k_0-1)}(x_0) = y_0^{(k_0-1)}; \\ H_n(x_1) = y_1; & H'_n(x_1) = y'_1; & \dots & H_n^{(k_1-1)}(x_1) = y_1^{(k_1-1)}; \\ \dots & \dots & \dots & \dots \\ H_n(x_m) = y_m; & H'_n(x_m) = y'_m; & \dots & H_n^{(k_m-1)}(x_m) = y_m^{(k_m-1)}. \end{array}$$

Единственность многочлена $H_n(x)$, удовлетворяющего условиям эрмитовой интерполяции, доказывается, как обычно, от противного. А именно, если $H_n(x)$ и $\tilde{H}_n(x)$ — два многочлена степени n , удовлетворяющие одним и тем же $n+1$ условиям (1.50), то это означает, что многочлен-разность $H_n(x) - \tilde{H}_n(x)$ степени не выше n имеет $n+1$ корень (с учетом кратностей), следовательно, $H_n(x) - \tilde{H}_n(x) \equiv 0$, отсюда — совпадение $H_n(x)$ и $\tilde{H}_n(x)$.

Существование многочлена $H_n(x)$, удовлетворяющего требованиям (1.50), можно вывести из его единственности. Действительно, возьмем в качестве исходной функции $y = f(x)$ функцию $y = 0$. Все ее значения и значения производных равны нулю, поэтому условия (1.50) для нее имеют вид

$$H_n^{(j)}(x_i) = 0, \quad (1.52)$$

и таких условий $n+1$. Получается, что многочлен n -й степени $H_n(x)$ имеет $n+1$ корень (с учетом кратностей); значит, это есть нуль-многочлен. А это означает, что все a_k в его выражении (1.51) равны нулю. С другой стороны, (1.52) — это однородная система линейных алгебраических уравнений относительно коэффициентов a_k , которая, в силу единственности такого набора a_k , имеет только тривиальное решение. Отсюда следует неравенство нулю ее детерминанта, влекущее разрешимость такой системы при любых правых частях, т.е. формальное существование наборов коэффициентов a_k ($k = 0, 1, \dots, n$), обеспечивающих выполнение условий эрмитовой интерполяции (1.50) для любой заданной с помощью (1.48) функции $y = f(x)$.

Выявление общего вида интерполяционных многочленов Эрмита $H_n(x)$ представляет непростую задачу и требует привлечения определенных сведений из теории функций комплексной переменной [67 и др.]. Рассмотрим одну из возможных процедур фактического построения таких многочленов, не требующую знания их общего вида (см., например, [11]; другой способ можно найти в [81]).

Пусть $L_m(x)$ — интерполяционный многочлен Лагранжа, построенный по данным $m+1$ значениям $y_i = f(x_i)$, $i = 0, 1, \dots, m$. Как и ранее (см.(1.10)), будем пользоваться обозначением $\Pi_{m+1}(x) := (x - x_0)(x - x_1) \dots (x - x_m)$. Так как по условию m заведомо не превосходит n , то по теореме о делении многочлена с остатком искомый многочлен Эрмита $H_n(x)$

можно представить в виде

$$H_n(x) = L_m(x) + H_{n-(m+1)} \cdot \Pi_{m+1}(x), \quad (1.53)$$

где $H_{n-(m+1)}(x)$ — некоторый неизвестный пока многочлен степени $n - m - 1$.

Для построения многочлена $H_{n-(m+1)}(x)$ будем привлекать информацию о производных данной функции, т.е. равенства $H'_n(x_i) = y'_i$ в тех узлах x_i , где первые производные, в соответствии с (1.48), заданы (информация о самих значениях функции уже полностью исчерпана: в силу $L_m(x_i) = y_i$ и $\Pi_{m+1}(x_i) = 0$ для всех x_i от x_0 до x_m , согласно (1.53), будет и $H_n(x_i) = y_i$ при любых $i \in \{0, 1, \dots, m\}$).

Продифференцировав равенство (1.53), имеем

$$H'_n(x) = L'_m(x) + H'_{n-(m+1)}(x) \cdot \Pi_{m+1}(x) + H_{n-(m+1)}(x) \cdot \Pi'_{m+1}(x). \quad (1.54)$$

Поскольку $\Pi_{m+1}(x_i) = 0$, в тех узлах x_i , где по условию эрмитовой интерполяции справедливо $H'_n(x_i) = y'_i$, можно записать

$$L'_m(x_i) + H_{n-(m+1)}(x_i) \cdot \Pi'_{m+1}(x_i) = y'_i.$$

Отсюда выражаем значения многочлена $H_{n-(m+1)}(x)$ в этих узлах:

$$H_{n-(m+1)}(x_i) = \frac{y'_i - L'_m(x_i)}{\Pi'_{m+1}(x_i)}.$$

Правая часть этого равенства может быть вычислена; обозначим ее через z'_i . Таким образом, в ряде узлов x_i известны значения многочлена $H_{n-(m+1)}(x_i) = z'_i$, по которым этот многочлен

однозначно восстанавливается обычной лагранжевой интерполяцией, если в условиях (1.48) не содержится производных порядка, выше первого (т.е. нет ни одного узла кратности больше 1); подстановка найденного многочлена $H_{n-(m+1)}(x)$ в (1.53) приводит к искомому интерполяционному многочлену Эрмита. Если же в исходной информации (1.48) об $f(x)$ имеются значения производных более высокого порядка, чем первый, то для восстановления многочлена $H_{n-(m+1)}(x)$ ставится задача эрмитовой интерполяции, для чего наряду с полученными его значениями z'_i , находят значения его производных путем дифференцирования равенства (1.54) (возможно неоднократно, в зависимости от максимального порядка заданных производных функции $f(x)$). Эта процедура построения интерполяционных многочленов Эрмита все более низких степеней продолжается до исчерпания всей информации (1.48) о функции и ее производных.

Рассмотрим реализацию описанного процесса эрмитовой интерполяции на простом примере, демонстрирующем возможность восстановления многочлена n -й степени по его значениям и значениям некоторых его производных при суммарной кратности узлов $n + 1$.

Пример 1.7. Пусть сведения о некоторой функции $y = f(x)$ представлены следующей дискретной информацией:

i	x_i	y_i	y'_i	y''_i
0	-1	0	-2	
1	0	1	0	-4
2	1	0	2	

В соответствии с обозначениями (1.48) здесь: $m = 2$; $k_0 - 1 = 1$, $k_1 - 1 = 2$, $k_2 - 1 = 1 \Rightarrow n + 1 = k_0 + k_1 + k_2 = 7 \Rightarrow n = 6$. Таким образом, по данным сведениям о функции $y = f(x)$, сосредоточенным в трех узлах $x_0 = -1$, $x_1 = 0$, $x_2 = 1$ кратности, соответственно, 2, 3, 2, следует строить интерполяционный многочлен Эрмита $H_6(x)$.

Согласно предложенной выше схеме, сначала, пользуясь столбцами x_i , y_i таблицы данных, записываем интерполяционный многочлен Лагранжа второй степени

$$L_2(x) = \frac{x(x-1)}{(-1)(-2)} \cdot 0 + \frac{(x+1)(x-1)}{1 \cdot (-1)} \cdot 1 + \frac{(x+1)x}{2 \cdot 1} \cdot 0 = 1 - x^2.$$

Далее по формуле (1.53) представляем $H_6(x)$ через $L_2(x)$, $H_3(x)$ и $\Pi_3(x)$:

$$H_6(x) = L_2(x) + H_3(x) \cdot \Pi_3(x) = 1 - x^2 + H_3(x)(x+1)x(x-1), \quad (1.55)$$

и дифференцируем этот многочлен дважды:

$$\begin{aligned} H'_6(x) &= -2x + (3x^2 - 1)H_3(x) + (x+1)x(x-1)H'_3(x); \\ H''_6(x) &= -2 + 6x \cdot H_3(x) + 2(3x^2 - 1)H'_3(x) + (x+1)x(x-1)H''_3(x). \end{aligned}$$

Подстановкой в $H'_6(x)$ значений $x = -1$, $x = 0$ и $x = 1$ и приравниванием $H'_6(x_i)$ заданным значениям y'_i получаем значения $H_3(x_i)$:

$$\text{при } x = -1 \quad H'_6(-1) = 2 + 2H_3(-1) = -2 \quad \Rightarrow \quad H_3(-1) = -2;$$

$$\text{при } x = 0 \quad H'_6(0) = 0 - H_3(0) = 0 \quad \Rightarrow \quad H_3(0) = 0;$$

$$\text{при } x = 1 \quad H'_6(1) = -2 + 2H_3(1) = 2 \quad \Rightarrow \quad H_3(1) = 2.$$

Учитывая их, из условия $H''_6(0) = -4$, т. е. из $-2 - 2H'_3(0) = -4$, находим $H'_3(0) = 1$.

Итак, для выявления многочлена $H_3(x)$ в (1.55) снова имеем задачу эрмитовой интерполяции с данными, содержащимися в следующей таблице:

i	x_i	$H_3(x_i)$	$H'_3(x_i)$
0	-1	-2	
1	0	0	1
2	1	2	

Здесь: $m = 2$; $k_0 = 1$, $k_1 = 2$, $k_2 = 1 \Rightarrow n = 3 \Rightarrow n - (m + 1) = 0$. В соответствии с (1.53), для этого случая записываем:

$$\begin{aligned} H_3(x) = \tilde{L}_2(x) + H_0 \cdot \Pi_3(x) &= \frac{x(x-1)}{(-1)(-2)}(-2) + \frac{(x+1)(x-1)}{1 \cdot (-1)} \cdot 0 + \frac{(x+1)x}{2 \cdot 1} \cdot 2 + \\ &+ H_0 \cdot (x+1)x(x-1) = 2x + x(x^2 - 1)H_0 \end{aligned}$$

(где $\tilde{L}_2(x)$ — многочлен Лагранжа, интерполирующий функцию $H_3(x)$).

Остается найти постоянную H_0 , для чего воспользуемся условием $H'_3(0) = 1$. Имеем:

$$H'_3(x) = 2 + (3x^2 - 1)H_0 \Rightarrow 2 + (3 \cdot 0 - 1)H_0 = 1 \Rightarrow H_0 = 1.$$

Следовательно,

$$H_3(x) = 2x + x(x^2 - 1) \cdot 1 = x^3 + x.$$

Подставив это в (1.55), получаем окончательное выражение искомого многочлена:

$$H_6(x) = 1 - x^2 + (x^3 + x)(x^3 - x) = x^6 - 2x^2 + 1.$$

Легко убедиться, что исходная таблица была составлена именно для этой функции.

В заключение отметим, что для $(n+1)$ -кратно дифференцируемой функции $f(x)$ *остаточный член интерполяционной формулы Эрмита* имеет вид [11, 85]

$$f(x) - H_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} (x-x_0)^{k_0} (x-x_1)^{k_1} \dots (x-x_m)^{k_m}, \quad (1.56)$$

где $k_0 + k_1 + \dots + k_m = n+1$ — суммарная кратность узлов $x_0, x_1, \dots, x_m$, а ξ — некоторая точка из промежутка $[a, b] \supseteq [x_0, x_m]$. Характерно, что в случае, когда все узлы — простые (однократные), т.е. $m=n$, интерполяционный многочлен Эрмита есть не что иное, как обычный многочлен Лагранжа $L_n(x)$, и остаточный член (1.56) совпадает с выведенным ранее его остаточным членом (1.12). Если же вся информация об $f(x)$ сосредоточена в одном узле x_0 , т.е. x_0 является узлом кратности $n+1$, то многочлен Эрмита — это просто многочлен Тейлора с

остаточным членом $\frac{f^{(n+1)}(\xi)}{(n+1)!} (x-x_0)^{n+1}$.

УПРАЖНЕНИЯ^{*}

1.1. Для функции $y = f(x)$, заданной тремя значениями $f(1) = 0.71$, $f(2) = 3.31$ и $f(3) = 0.18$, найдите коэффициенты интерполирующего ее многочлена $P_2(x) = a_0 + a_1x + a_2x^2$ непосредственно из условий интерполяции.

1.2. Привлекая интерполяционные соображения, докажите, что выражение

$$\frac{(x-b)(x-c)}{(a-b)(a-c)} + \frac{(x-a)(x-c)}{(b-a)(b-c)} + \frac{(x-a)(x-b)}{(c-a)(c-b)}$$

тождественно равно 1 при любых попарно различающихся a, b и c .

1.3. Дана таблица значений функции $y = \lg x$:

x_i	11	12	13	14	15
y_i	1.0414	1.0792	1.1139	1.1461	1.1761

а) С помощью линейной интерполяционной формулы Лагранжа вычислите $\lg 11.6$ ($\approx L_1(11.6)$), оцените погрешность и сравните ее с фактической ошибкой.

^{*} Предполагается, что все фигурирующие в упражнениях функции обладают нужной гладкостью.

б) С какой точностью можно вычислить по этим данным $\lg 11.6$ посредством интерполяционной формулы Лагранжа третьей степени? Запишите расчетную формулу для вычисления $\lg 11.6 \approx L_3(11.6)$.

в) Можно ли в данных условиях построить интерполяционный многочлен пятой степени?

1.4. Пользуясь интерполяционной схемой Эйткена, пополните таблицу

x	2.0	2.1	2.2	2.3	2.4	2.5
y	0.3010		0.3424		0.3802	0.3979

недостающими значениями $y(2.1)$ и $y(2.3)$.

1.5. Какую точность можно гарантировать при линейной интерполяции функции $y = xe^{-x}$ на отрезке $[0, 1]$, считая узлами интерполяции его концы?

1.6. Восстановите многочлен $P_3(x)$ по его значениям:

x	-1	0	1	2
$P_3(x)$	1	1	-1	7

1.7. По данным упражнения 1.3 постройте таблицу конечных разностей. Вычислите по интерполяционным формулам Ньютона приближенные значения $\lg 11.6$, $\lg 10.5$, $\lg 14.5$ и $\lg 15.2$. Сравните их с точными значениями.

1.8. Применяя наиболее подходящие центральные интерполяционные формулы, найдите значения $y(1)$, $y(1.12)$ и $y(1.5)$, если соответствие между x и $y = y(x)$ задано таблицей

x	0.1	0.5	0.9	1.3	1.7	2.1
y	0.000	0.100	0.354	0.734	1.189	1.676

1.9. В каких точках выполняется условие лагранжевой интерполяции:

а) для многочлена Бесселя второй степени, построенного по точкам x_{-1}, x_0, x_1, x_2 ?

б) для многочлена Стирлинга третьей степени, построенного по точкам $x_{-2}, x_{-1}, x_0, x_1, x_2$?

1.10. По аналогии с формулой (1.43) выведите вторую интерполяционную формулу Ньютона, пригодную для интерполирования по неравным промежуткам в конце таблицы значений сеточной функции.

1.11. Для функции $y = f(x)$, заданной таблицей

x	0.5	0.7	1.0	1.4	2.0	2.6	4.0
$f(x)$	-0.555	-0.239	0.000	0.114	0.139	0.123	0.082

составьте таблицу разделенных разностей, запишите подходящие для вычисления $f(0.6)$, $f(1.5)$ и $f(3)$ конкретные интерполяционные многочлены и найдите эти приближенные значения.

1.12. Докажите, что в примере 1.6 (см. § 1.7) выполняются достаточные условия сходимости итерационного процесса (1.47).

1.13. Некоторая зависимость $y = y(x)$ задана следующей таблицей:

x	1	7	13	19	25
y	0.21361	0.40541	0.85225	1.79725	3.78183

Обратным интерполированием установите, каким значениям x отвечают значения $y = 0.3178$ и $y = 1.159$.

1.14. Найдите многочлен $P(x)$, данные о котором представлены следующей таблицей:

x	$P(x)$	$P'(x)$	$P''(x)$
-1	15	-14	-2
0	4	-7	
2	18		

ГЛАВА 2 || МНОГОЧЛЕНЫ ЧЕБЫШЕВА И || НАИЛУЧШИЕ РАВНОМЕРНЫЕ || ПРИБЛИЖЕНИЯ

Дается определение многочленов Чебышева первого рода и изучаются их свойства. Показывается, что погрешность интерполирования гладкой функции многочленом фиксированной степени будет наименьшей, когда в качестве узлов интерполяции используются корни многочленов Чебышева. Приводится (без доказательства) теорема о чебышевском альтернансе, служащая теоретической основой построения наилучших равномерных приближений, и рассматриваются простейшие ситуации, когда такие многочлены могут быть построены. Обсуждается идея использования разложения функций в степенные ряды для получения многочленов, близких к многочленам наилучших равномерных приближений.

§ 2.1. ОПРЕДЕЛЕНИЕ И СВОЙСТВА МНОГОЧЛЕНОВ ЧЕБЫШЕВА

Определение 2.1. *Многочленом Чебышева^{*)} называется функция*

$$T_n(x) := \cos(n \arccos x), \quad (2.1)$$

где $n \in \mathbb{N}_0$, $x \in [-1, 1]$.

Прежде всего, убедимся, что функция $T_n(x)$, представленная с помощью тригонометрических функций, на самом деле является многочленом при любом $n = 0, 1, 2, \dots$

Непосредственной подстановкой в (2.1) значений $n = 0$ и $n = 1$ получаем $T_0(x) = 1$, $T_1(x) = x$.

Положив $\alpha := \arccos x$, имеем:

$$T_1(x) = \cos \alpha, \quad T_n(x) = \cos n\alpha,$$

$$T_{n-1}(x) = \cos(n-1)\alpha, \quad T_{n+1}(x) = \cos(n+1)\alpha,$$

^{*)} Чебышёв Пафнутий Львович (1821–1894) — знаменитый русский математик. Его работы о многочленах наилучшего равномерного приближения легли в основу конструктивной теории функций. Стандартное обозначение $T_n(x)$ можно соотнести с немецким написанием фамилии Tschebyschew.

и так как (по формуле суммы косинусов)

$$\cos(n+1)\alpha + \cos(n-1)\alpha = 2\cos\alpha\cos n\alpha,$$

то, значит, справедливо равенство

$$T_{n+1}(x) + T_{n-1}(x) = 2T_1(x)T_n(x),$$

которое может быть переписано в виде

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x). \quad (2.2)$$

Формула (2.2) определяет при $n = 1, 2, 3, \dots$ последовательность функций $T_n(x)$, начинающуюся с $T_0(x) = 1$, $T_1(x) = x$, рекуррентно; при этом нужно иметь в виду, что здесь $x \in [-1, 1]$, как и в (2.1).

Подставляя в (2.2) заданные начальные члены последовательности ($T_n(x)$), найдем несколько ее последующих членов:

$$T_2(x) = 2x^2 - 1;$$

$$T_3(x) = 2x(2x^2 - 1) - x = 4x^3 - 3x;$$

$$T_4(x) = 2x(4x^3 - 3x) - 2x^2 + 1 = 8x^4 - 8x^2 + 1;$$

$$T_5(x) = 2x(8x^4 - 8x^2 + 1) - 4x^3 + 3x = 16x^5 - 20x^3 + 5x$$

и т. д.

Графики нескольких многочленов Чебышева (с первого по четвертый) изображены на рис. 2.1.

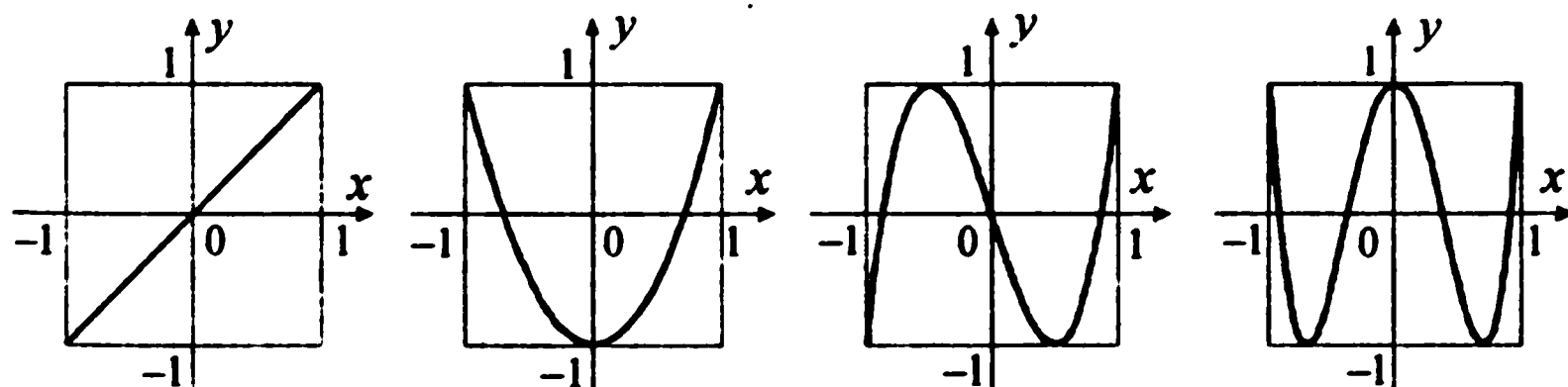


Рис. 2.1. Графики многочленов $T_1(x)$, $T_2(x)$, $T_3(x)$ и $T_4(x)$

Анализ рекуррентной формулы (2.2) позволяет считать очевидными следующие факты:

1) все функции $T_n(x)$, определенные в (2.1), являются многочленами при любом натуральном n ;

2) степени этих многочленов возрастают с увеличением n , причем старший член многочлена $T_n(x)$ равен $2^{n-1}x^n$;

3) многочлены $T_n(x)$ при четных n выражаются через степенные функции только четных степеней, при нечетных — только нечетных.

Наряду с многочленами Чебышева $T_n(x)$, часто используют многочлены, получаемые из $T_n(x)$ делением на старший коэффициент, т.е.

$$\hat{T}_n(x) := \frac{1}{2^{n-1}} T_n(x)$$

— многочлены со старшим коэффициентом 1. Будем называть их **нормированными многочленами Чебышева**.

Многочлены Чебышева обладают рядом замечательных свойств. Рассмотрим некоторые их свойства, имеющие отношение к поставленной выше проблеме аппроксимации функций.

Свойство 2.1. Многочлен Чебышева $T_n(x)$ (а значит, и многочлен $\hat{T}_n(x)$) имеет на отрезке $[-1, 1]$ ровно n различных действительных корней; все они задаются формулой

$$x_k = \cos \frac{2k+1}{2n} \pi, \text{ где } k = 0, 1, \dots, n-1. \quad (2.3)$$

Доказательство. Корни многочлена Чебышева $T_n(x)$ можно получить, решая тригонометрическое уравнение

$$\cos(n \arccos x) = 0.$$

Имеем:

$$n \arccos x = \frac{\pi}{2} + \pi k \Leftrightarrow \arccos x = \frac{2k+1}{2n} \pi.$$

Поскольку главные значения арккосинуса должны принадлежать промежутку $[0, \pi]$, целая переменная k здесь должна принимать значения от 0 до $n-1$. Переходя на этом промежутке к обратной функции, иначе, беря косинус от левой и правой частей последнего равенства и учитывая, что $\cos \arccos x = x$, приходим к выводу, что n действительных корней многочлена $T_n(x)$, определяемые формулой (2.3), на самом деле существуют. Их несовпадение и принадлежность отрезку $[-1, 1]$ следует из того, что они суть значения косинуса, соответствующие различным значениям его аргумента, расположенным на промежутке его монотонности.

Свойство 2.2. *Корни многочленов Чебышева перемежаются с точками их наибольших и наименьших значений, равных соответственно $+1$ и -1 для $T_n(x)$ и $+\frac{1}{2^{n-1}}$ и $-\frac{1}{2^{n-1}}$ для $\hat{T}_n(x)$. А именно, при $j = 0, 1, \dots, n$ имеют место экстремумы*

$$T_n(x_j) = (-1)^j, \quad \hat{T}_n(x_j) = \frac{(-1)^j}{2^{n-1}} \text{ в точках } x_j = \cos \frac{j}{n} \pi. \quad (2.4)$$

Доказательство. Из определения $T_n(x)$ формулой (2.1) следует, что $|T_n(x)| \leq 1$. Покажем, что максимальное значение 1 для $|T_n(x)|$ достигается, т.е. существуют такие точки $x \in [-1, 1]$, в которых $|T_n(x)| = 1$. Имеем:

$$T_n(x) = \pm 1 \Leftrightarrow \cos(n \arccos x) = \pm 1 \Leftrightarrow$$

$$n \arccos x = \pi j \quad (j \in \mathbb{Z}) \Leftrightarrow \arccos x = \frac{j}{n} \pi \quad (j \in \mathbb{Z}).$$

Число $\frac{j}{n} \pi$ принадлежит отрезку $[0, \pi]$, т.е. может служить значением арккосинуса, если $j \in \{0, 1, \dots, n\}$. Следовательно, уравнение $|T_n(x)| = 1$ имеет своими корнями $n+1$ точек

$$x_j = \cos \frac{j}{n} \pi, \quad j = 0, 1, \dots, n. \quad (2.5)$$

Что эти точки экстремумов $T_n(x)$ перемежаются с корнями $\hat{T}_n(x)$, легко увидеть, переписав (2.5) в виде $x_j = \cos \frac{2j}{2n} \pi$ и сравнив это с (2.3). Факт чередования максимумов и минимумов у $T_n(x)$ следует из того, что значения выражения $n \arccos x$ в точках x_j , согласно промежуточному равенству $n \arccos x = \pi j$, равны поочередно $0, \pi, 2\pi, 3\pi, \dots$, и потому, в зависимости от четности j , значения $\cos(n \arccos x)$ будут соответственно $+1$, -1 .

Свойство 2.3 (теорема Чебышева). *Из всех многочленов степени n со старшим коэффициентом 1 нормированный многочлен Чебышева $\hat{T}_n(x)$ наименее уклоняется от нуля на отрезке $[-1, 1]$.*

Доказательство (от противного). Пусть существует многочлен $\tilde{P}_n(x) = x^n + \tilde{a}_1 x^{n-1} + \dots + \tilde{a}_n$ такой, что

$$\max_{x \in [-1, 1]} |\tilde{P}_n(x)| < \max_{x \in [-1, 1]} |\hat{T}_n(x)| \quad \left(= \frac{1}{2^{n-1}} \text{ по свойству 2.2} \right).$$

Разность этих двух многочленов, очевидно, есть многочлен степени, не выше $n-1$; обозначим его

$$Q_{n-1}(x) := \hat{T}_n(x) - \tilde{P}_n(x) = b_1 x^{n-1} + b_2 x^{n-2} + \dots + b_n.$$

Пользуясь тем, что график $\hat{T}_n(x)$ лежит в полосе, ограниченной прямыми $y = \pm \frac{1}{2^{n-1}}$, чередуя касания то верхней, то нижней из этих прямых (по свойству 2.2), а график $\tilde{P}_n(x)$ при $x \in [-1, 1]$ должен лежать строго внутри этой полосы (по предположению), то можно утверждать, что многочлен-разность $Q_{n-1}(x)$ в точках экстремумов $x_j = \cos \frac{j}{n} \pi$ должен иметь определенные знаки: $+$ при четных j и $-$ при нечетных j (рис. 2.2).

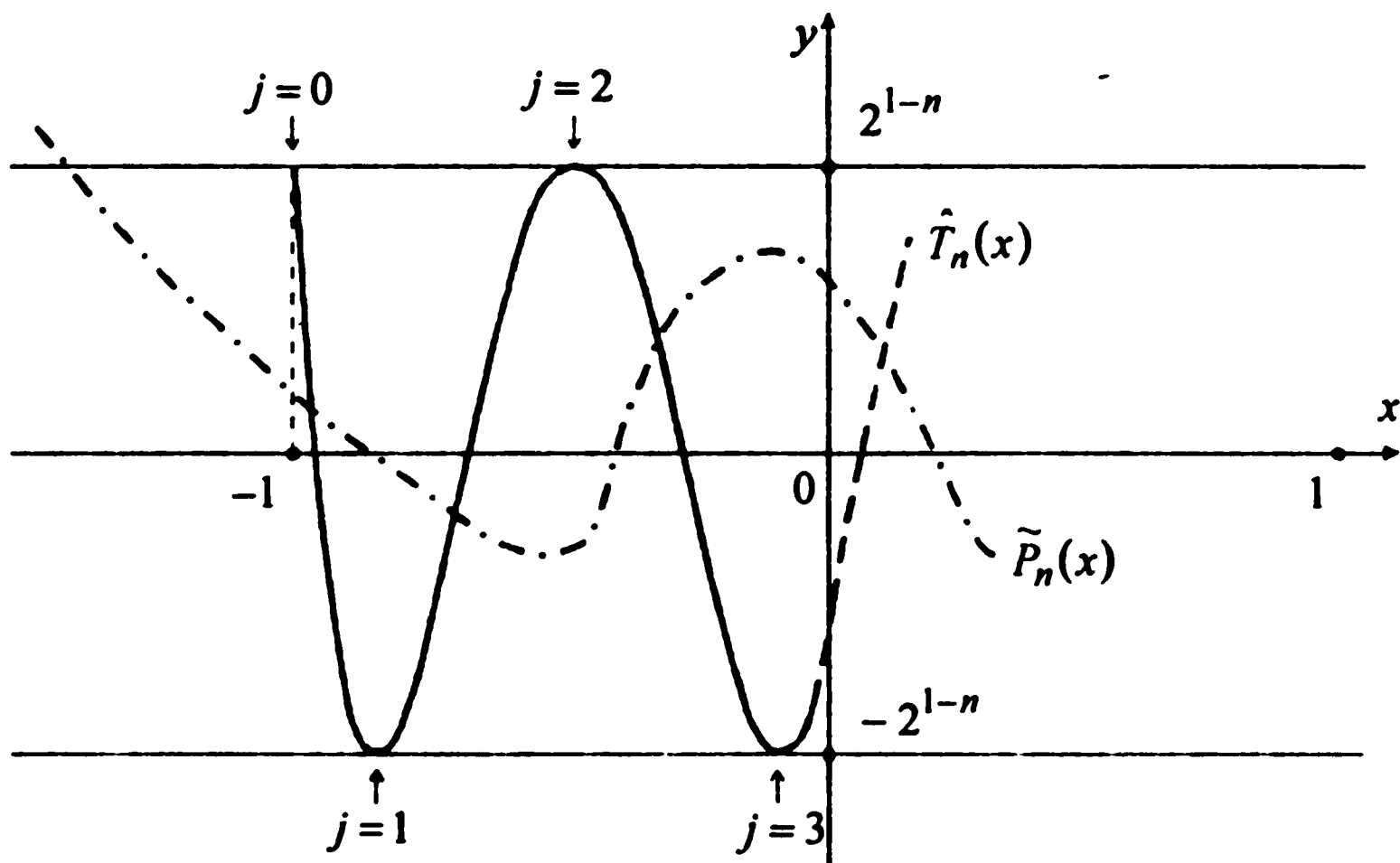


Рис. 2.2. Возможное совместное поведение графиков $\hat{T}_n(x)$ и $\tilde{P}_n(x)$

Поскольку $\hat{T}_n(x)$ на $[-1, 1]$ имеет $n+1$ таких точек экстремумов, следовательно, многочлен $Q_{n-1}(x)$ должен иметь, по меньшей

мере, n перемен знаков, т.е. n корней, что противоречит следствию из основной теоремы алгебры многочленов. Полученное противоречие говорит о несостоятельности сделанного в начале доказательства предположения.

Только что доказанное свойство означает, что *среди всех многочленов степени n вида*

$$P_n(x) = x^n + a_1 x^{n-1} + \dots + a_n \quad (2.6)$$

именно нормированный многочлен Чебышева $\hat{T}_n(x)$ минимизирует максимальное расстояние от графика многочлена при $x \in [-1, 1]$ до оси абсцисс, т.е. $\hat{T}_n(x)$ — это многочлен с наименьшей нормой (так называемой чебышевской нормой) на множестве многочленов вида (2.6) в пространстве $C[-1, 1]$.

Замечание 2.1. Иногда бывает более удобным использовать *смещенные многочлены Чебышева $T_n^*(x)$* , которые определяются на отрезке $[0, 1]$ и могут быть получены из классических многочленов Чебышева $T_n(x)$ заменой в них аргумента x на аргумент $2x - 1$.

§ 2.2. ИНТЕРПОЛЯЦИЯ ПО ЧЕБЫШЕВСКИМ УЗЛАМ

Вернемся к изучавшейся в предыдущей главе задаче интерполяции.

Сравнивая конечноразностные интерполяционные многочлены, построенные по системе равноотстоящих узлов, с интерполяционным многочленом Лагранжа, предполагающим произвольное расположение несовпадающих узлов на промежутке интерполирования $[a, b]$, следует отметить, что первые более просты и удобны в использовании, вторые же обладают большими возможностями. Нет сомнений в том, что если можно располагать узлы в пределах отрезка $[a, b]$ как угодно, то имеет смысл использовать большее количество точечной информации о функции там, где она более сильно изменяется. Особенно существенным это замечание может оказаться при эрмитовой интерполяции.

Подойдем к проблеме расположения узлов интерполяции с несколько иной стороны.

Желая, чтобы интерполяционный многочлен Лагранжа $L_n(x)$ (1.6) в целом хорошо приближал функцию $y = f(x)$ на отрезке $[a, b]$, поставим вопрос: как расположить на нем $n + 1$

узлов интерполяции x_i ($i = 0, 1, \dots, n$), чтобы при этом минимизировать максимальную на $[a, b]$ погрешность?

Поскольку, согласно (1.15),

$$\max_{x \in [a, b]} |f(x) - L_n(x)| \leq \frac{M_{n+1}}{(n+1)!} \max_{x \in [a, b]} |\Pi_{n+1}(x)|, \quad (2.7)$$

и величина $M_{n+1} := \max_{x \in [a, b]} |f^{(n+1)}(x)|$ не зависит от расположения узлов, займемся изучением определенного в (1.10) многочлена $\Pi_{n+1}(x)$.

Сначала выполним одно простое преобразование независимой переменной, которое неоднократно будет использоваться и в дальнейшем. А именно, положим

$$x = \frac{a+b}{2} + \frac{b-a}{2}t. \quad (2.8)$$

Легко видеть, что эта линейная замена осуществляет взаимно-однозначное соответствие между $x \in [a, b]$ и $t \in [-1, 1]$.

Благодаря (2.8) узлам $x_i \in [a, b]$ можно сопоставить точки $t_i \in [-1, 1]$, полагая

$$x_i = \frac{a+b}{2} + \frac{b-a}{2}t_i.$$

Тогда образующие $\Pi_{n+1}(x)$ сомножители $x - x_i$ преобразуются по формуле

$$x - x_i = \frac{b-a}{2}(t - t_i),$$

и, следовательно,

$$\begin{aligned} \Pi_{n+1}(x) &:= \prod_{i=0}^n (x - x_i) = \left(\frac{b-a}{2}\right)^{n+1} \prod_{i=0}^n (t - t_i) = \\ &= \left(\frac{b-a}{2}\right)^{n+1} (t - t_0)(t - t_1)\dots(t - t_n). \end{aligned} \quad (2.9)$$

Из (2.9) явствует, что значение $\max_{x \in [a, b]} |\Pi_{n+1}(x)|$ будет минимальным, когда будет минимальным значение $\max_{t \in [-1, 1]} |\Pi_{n+1}(t)|$, где

$$\Pi_{n+1}(t) := (t - t_0)(t - t_1)\dots(t - t_n)$$

— многочлен $n+1$ степени, в котором точки $t_i \in [-1, 1]$ ($i = 0, 1, \dots, n$) считаем параметрами. Так как коэффициент при

старшей степени многочлена $\Pi_{n+1}(t)$ в его каноническом представлении равен 1, то, в силу свойства 2.3, величина $\max_{t \in [-1, 1]} |\Pi_{n+1}(t)|$ (а значит, и $\max_{x \in [a, b]} |\Pi_{n+1}(x)|$) будет минимальной в том случае, когда $\Pi_{n+1}(t)$ есть нормированный многочлен Чебышева $\hat{T}_{n+1}(t)$.

Таким образом, полагая $\Pi_{n+1}(t) = \hat{T}_{n+1}(t)$, имеем: с одной стороны, точки $t_0, t_1, \dots, t_n$ по замыслу являются узлами интерполяции (для переведенной задачи интерполяции с исходного отрезка $[a, b]$ на отрезок $[-1, 1]$), с другой стороны, в силу структуры многочлена $\Pi_{n+1}(t)$, они служат его корнями, а значит, и корнями $\hat{T}_{n+1}(t)$. Отсюда — вывод:

*максимальная погрешность интерполирования достаточно гладкой функции на отрезке $[-1, 1]$ многочленом n -й степени будет минимальной, когда в качестве узлов интерполяции $t_0, t_1, \dots, t_n \in [-1, 1]$ берутся корни многочлена Чебышева $T_{n+1}(t)$ (как известно, совпадающие с корнями $\hat{T}_{n+1}(t)$). Будем называть их **чебышевскими узлами интерполяции**.*

Знание экстремальных значений многочлена Чебышева (см. (2.4)) позволяет уточнить величину максимального отклонения $L_n(x)$ от $f(x)$ при таком выборе узлов, т.е. когда точки t_i ($= \cos \frac{2i+1}{2n+2} \pi$, $i = \overline{0, n}$) есть корни $\hat{T}_{n+1}(t)$, а $x_i = \frac{a+b}{2} + \frac{b-a}{2} t_i$.

А именно, так как $\max_{t \in [-1, 1]} |\hat{T}_{n+1}(t)| = \frac{1}{2^n}$, то, согласно (2.9),

$$\max_{x \in [a, b]} |\Pi_{n+1}(t)(x)| = \left(\frac{b-a}{2} \right)^{n+1} \cdot \frac{1}{2^n} = \frac{(b-a)^{n+1}}{2^{2n+1}};$$

подставляя это в неравенство (2.7), получаем

$$\max_{x \in [a, b]} |f(x) - L_n(x)| \leq \frac{M_{n+1}}{(n+1)!} \frac{(b-a)^{n+1}}{2^{2n+1}}. \quad (2.10)$$

Найденная оценка (2.10) называется **наилучшей равномерной оценкой погрешности интерполяции**.

Покажем ее **неулучшаемость**, т.е. что существуют такие функции $f(x)$, для которых нестрогое неравенство (2.10) реализуется в виде равенства. С этой целью будем рассматривать аппроксимацию многочленом Лагранжа n -й степени, построенным

по чебышевским узлам интерполяции, функции $f(x)$, представляющей собой некоторый, вообще говоря, произвольный многочлен $n + 1$ -й степени

$$P_{n+1}(x) := a_0 x^{n+1} + a_1 x^n + a_2 x^{n-1} + \dots + a_n x + a_{n+1}.$$

Так как для такой функции $f(x)$ производная $n + 1$ -го порядка есть $f^{(n+1)}(x) \equiv P_{n+1}^{(n+1)}(x) \equiv a_0(n+1)!$, то в неравенстве (2.10) можно считать $M_{n+1} = a_0(n+1)!$, и сама эта оценка (2.10) превращается в оценку

$$\max_{x \in [a, b]} |P_{n+1}(x) - L_n(x)| \leq a_0 \frac{(b-a)^{n+1}}{2^{2n+1}}.$$

Рассматривая же фактическую разность между $P_{n+1}(x)$ и $L_n(x)$, в силу постоянства $n + 1$ -й производной, т.е. независимости $P_{n+1}(\xi)$ от положения точки ξ на (a, b) , имеем:

$$P_{n+1}(x) - L_n(x) = \frac{P_{n+1}^{(n+1)}(\xi)}{(n+1)!} \Pi_{n+1}(x) = a_0 \left(\frac{b-a}{2} \right)^{n+1} \hat{T}_{n+1}(t) \quad (2.11)$$

(см. (1.13) и (2.9); в последнем случае следует принять во внимание, что $\Pi_{n+1}(t) = \hat{T}_{n+1}(t)$ в соответствии со спецификой фиксирования точек t_i). Зная точное значение $\max_{t \in [-1, 1]} |\hat{T}_{n+1}(t)|$, на основе (2.11) получаем:

$$\max_{x \in [a, b]} |P_{n+1}(x) - L_n(x)| = a_0 \left(\frac{b-a}{2} \right)^{n+1} \frac{1}{2^n} = a_0 \frac{(b-a)^{n+1}}{2^{2n+1}}.$$

Как видим, фактическая максимальная разность и ее оценка по формуле (2.10) на функции $f(x) = P_{n+1}(x)$ совпали, значит оценка (2.10) достижима, т.е. правая часть ее в общем случае не может быть уменьшена.

§ 2.3. О МНОГОЧЛЕНАХ НАИЛУЧШИХ РАВНОМЕРНЫХ ПРИБЛИЖЕНИЙ

Резюмируя исследования предыдущего параграфа, в частности, анализируя неравенство (2.10), можно хотя бы частично ответить на вопрос о сходимости последовательности интерполяционных многочленов Лагранжа $L_n(x)$ к функции $f(x)$, для которой они построены:

если функция $f(x)$ бесконечно дифференцируема на $[a, b]$ и в качестве узлов интерполяции берутся корни многочленов Чебышева (приведенные к отрезку $[a, b]$), то

$$\max_{x \in [a, b]} |f(x) - L_n(x)| \xrightarrow{n \rightarrow \infty} 0.$$

Поскольку $\max_{x \in [a, b]} |f(x) - L_n(x)|$ есть не что иное, как $\|f(x) - L_n(x)\|_{C[a, b]}$ — **чебышевская норма** в пространстве $C[a, b]$ непрерывных на $[a, b]$ функций, можно говорить, что для достаточно гладких функций $f(x)$ при специальном расположении узлов интерполяции последовательность интерполяционных многочленов Лагранжа ($L_n(x)$) (построенных по точным значениям функции $f(x)$) сходится к $f(x)$ по норме Чебышева; другими словами, имеет место равномерная сходимость последовательности ($L_n(x)$) к $f(x)$.

Обобщением установленного факта для непрерывных (не обязательно дифференцируемых) функций и произвольных (не обязательно интерполяционных) многочленов является широко известная в математическом анализе теорема.

Теорема 2.1 (Вейерштрасса) [5, 24, 43, 58, 75]. Для любой непрерывной на $[a, b]$ функции $f(x)$ найдется многочлен $Q_n(x)$ такой, что

$$|f(x) - Q_n(x)| < \varepsilon \quad \forall \varepsilon > 0 \quad \forall x \in [a, b].$$

Как следует из теоремы Вейерштрасса, если отказаться от того, чтобы аппроксимирующий функцию $f(x)$ многочлен $Q_n(x)$ степени n был интерполяционным, от $f(x)$ достаточно потребовать непрерывность на $[a, b]$, чтобы за счет повышения степени многочлена при надлежащем подборе его коэффициентов величина чебышевской нормы $\|f(x) - Q_n(x)\|_{C[a, b]}$ могла быть сделанной сколь угодно малой, иначе, чтобы качество аппроксимации функции $f(x)$ многочленом $Q_n(x)$ на отрезке $[a, b]$ было сколь угодно хорошим в смысле чебышевской метрики.

Если степень n аппроксимирующего $f(x)$ многочлена $Q_n(x)$ зафиксировать и распоряжаться только его коэффициентами, то в общем случае величину $\|f(x) - Q_n(x)\|_{C[a, b]}$ нельзя

сделать сколь угодно малой. Однако доказано [23], что для любой функции $f(x) \in C[a, b]$ существует единственный многочлен $Q_n^f(x)$ такой, который из всех многочленов $Q_n(x)$ степени n наилучшим образом аппроксимирует на $[a, b]$ функцию $f(x)$, минимизируя максимальное расстояние между $f(x)$ и $Q_n(x)$. Этот многочлен, т.е. многочлен $Q_n^f(x)$ такой, что

$$\|f(x) - Q_n^f(x)\|_{C[a, b]} = \inf \|f(x) - Q_n(x)\|_{C[a, b]},$$

называется **многочленом наилучшего равномерного приближения** для $f(x)$ на $[a, b]$ или ее **чебышевским приближением**^{*)}.

Для последовательности величин

$$E_n(f) := \|f(x) - Q_n^f(x)\|_{C[a, b]},$$

представляющих собой погрешности аппроксимации функции $f(x) \in C[a, b]$ посредством многочлена $Q_n^f(x)$, в соответствии с теоремой Вейерштрасса можно установить монотонную сходимость к нулю, т. е. что

$$E_n(f) \leq E_{n-1}(f) \quad \text{и} \quad E_n(f) \xrightarrow{n \rightarrow \infty} 0.$$

Одним из характеристических свойств многочленов наилучших равномерных приближений является **критерий Чебышева**. Его отражает следующая теорема.

Теорема 2.2 (Чебышева). *Многочлен $Q_n^f(x)$ является многочленом наилучшего равномерного приближения для функции $f(x) \in C[a, b]$ тогда и только тогда, когда на $[a, b]$ существует не менее $n + 2$ точек x_i таких, что в них поочередно принимаются наибольшие положительные и отрицательные отклонения, т. е. поочередно разность $f(x_i) - Q_n^f(x_i)$ равна E или $-E$, где*

$$E := \|f(x) - Q_n^f(x)\|_{C[a, b]} = \max_{x \in [a, b]} |f(x) - Q_n^f(x)|.$$

Эта теорема (доказательство которой можно найти, например, в [5, 9]) говорит о том, что максимальная ошибка аппроксимации функции многочленом наилучшего равномерного приближения реализуется в числе точек, большем, по меньшей мере, на 2, чем степень многочлена, причем знаки ошибки чередуются. Точки x_i , в которых реализуется максимальное отклонение

^{*)} Первое доказательство существования многочлена наилучшего равномерного приближения для произвольной непрерывной функции дано французским математиком Э. Борелем (1871–1956).

многочлена $Q_n^f(x)$ от $f(x)$ на $[a, b]$, называются **точками чебышевского альтернанса**.

К сожалению, неизвестны ни общий вид многочленов наилучших равномерных приближений, ни способы их построения. Имеются лишь некоторые методики построения многочленов, близких к наилучшим равномерным (см., например, [32, 71, 78]), а также способы построения чебышевских приближений невысокого порядка для нескольких весьма узких классов функций. Последние существенно опираются на приведенную теорему о чебышевском альтернансе, что демонстрируется в следующих двух простейших случаях.

Случай А. Пусть функция $f(x)$ непрерывна на $[a, b]$, и пусть для нее требуется построить многочлен наилучшего равномерного приближения нулевой степени. Обозначим этот приближающий многочлен через $\varphi_0(x)$. Он определяется всего одним параметром: $\varphi_0(x) = A_0$. Чтобы найти значение этого параметра для заданной функции $f(x)$, воспользуемся тем свойством непрерывной на замкнутом промежутке функции, согласно которому на нем всегда найдутся, по крайней мере, две точки, в которых она принимает свои наименьшее и наибольшее значения.

Пусть $\min_{x \in [a, b]} f(x) = m$, $\max_{x \in [a, b]} f(x) = M$. Тогда совершенно очевидно, что полагая $A_0 = \frac{m + M}{2}$, т. е. подменяя функцию $f(x)$ функцией $\varphi_0 = \frac{m + M}{2}$, будем иметь максимальное отклонение

$$\begin{aligned} E = \|f(x) - \varphi_0(x)\|_{C[a, b]} &= \max_{x \in [a, b]} \left| f(x) - \frac{m + M}{2} \right| = \\ &= M - \frac{m + M}{2} = \frac{m + M}{2} - m = \frac{M - m}{2}, \end{aligned}$$

причем точки отрезка $[a, b]$, в которых оно реализуется — это точки, где принимаются значения m и M . В силу непрерывности $f(x)$, локальные минимумы и максимумы должны чередоваться; по меньшей мере, два из них определяют точки чебышевского альтернанса (с чередованием знаков разностей $f(x) - \frac{m + M}{2}$). Поэтому не существует другой постоянной, которая приближала бы $f(x)$ на $[a, b]$ лучше, чем постоян-

ная $\frac{m+M}{2}$, в смысле чебышевской метрики (на рис. 2.3 в качестве чебышевского альтернанса могут служить пары точек x_1, x_2 или x_1, x_3).

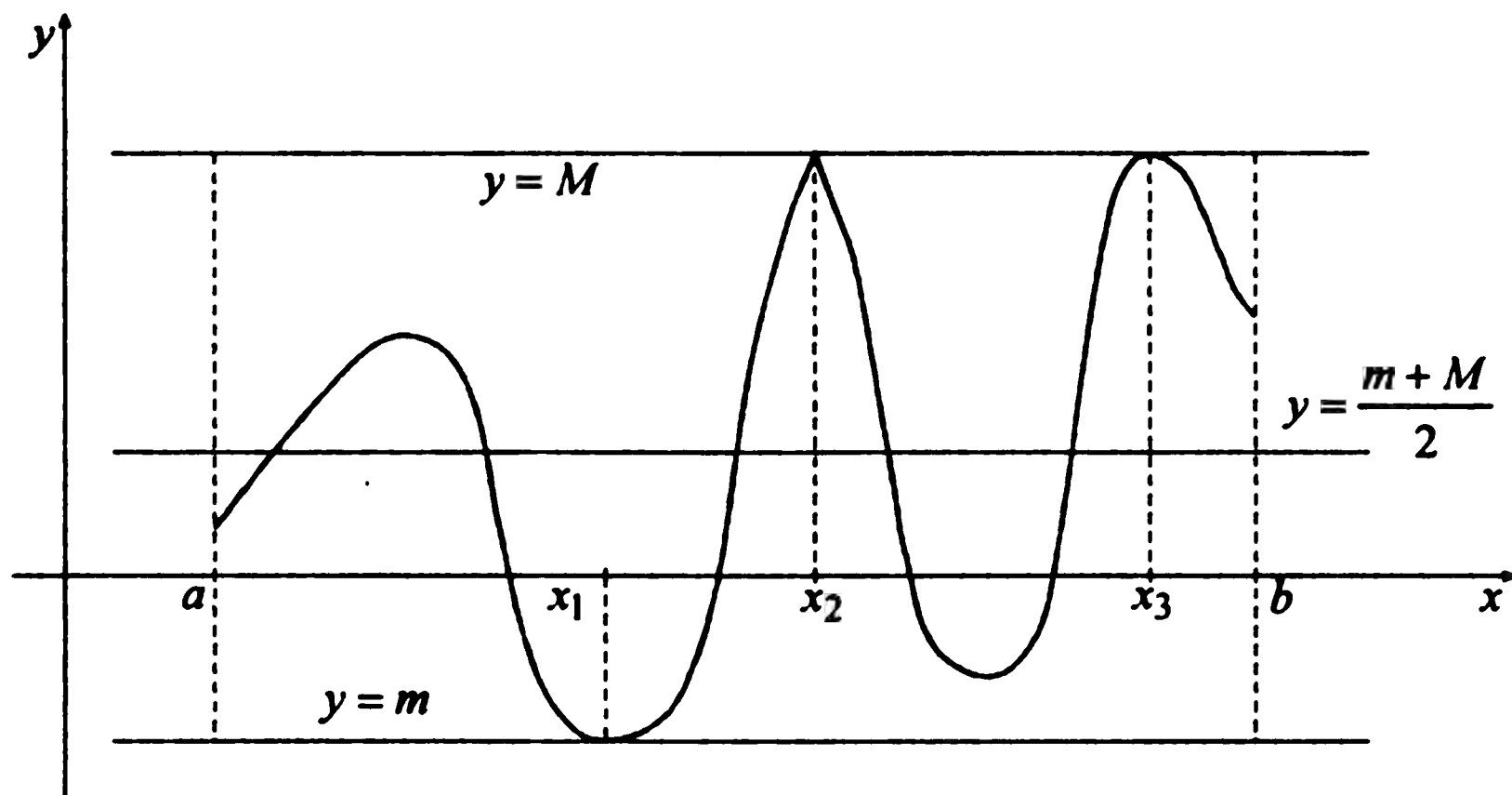


Рис. 2.3. Наилучшее равномерное приближение функции $f(x)$ с помощью постоянной

Случай Б. Пусть аппроксимируемая функция $f(x)$ дифференцируема и выпукла (в широком смысле) на отрезке $[a, b]$, а аппроксимирующая ее функция $\varphi(x) = A_0 + A_1x$ — многочлен наилучшего равномерного приближения первой степени. Чтобы найти его коэффициенты A_0 и A_1 , следует изучить разность между $f(x)$ и $\varphi(x)$, т. е. функцию $u(x) := f(x) - A_0 - A_1x$.

Так как функция $f(x)$ по предположению выпукла, а сдвиг на линейную функцию $A_0 + A_1x$ не изменяет выпуклости, то и функция $u(x)$ выпукла на $[a, b]$. Если речь идет о выпуклости вниз, то, как известно из анализа выпуклых функций, выпуклость $u(x)$ на $[a, b]$ влечет ее унимодальность на этом отрезке, т. е. существует единственная точка $c \in [a, b]$, в которой $u(x)$ имеет минимум; если $u(x)$ выпукла вверх, то в точке c должен быть максимум $u(x)$. В любом случае, точка $c \in [a, b]$ есть точка экстремума $u(x)$, и за счет возможности варьирования коэффициентов функции $\varphi(x)$ (точнее, коэффициента A_1) можно считать, что точка c является внутренней точкой отрезка $[a, b]$.

Потребуем, чтобы точки a, c и b в указанной последовательности составляли чебышевский альтернанс, т. е. чтобы в них последовательно принимались значения $E, -E, E$ или $-E, E,$

— E , где $E := \|f(x) - \varphi(x)\|_{C[a,b]} = \max_{x \in [a,b]} |f(x) - \varphi(x)|$ — максимальная погрешность аппроксимации функции $f(x)$ функцией $\varphi(x)$. Добавляя к этим требованиям еще необходимое условие экстремума дифференцируемой функции $u(x)$ в точке c и возвращаясь к исходным функциям, приходим к системе четырех уравнений относительно четырех неизвестных A_0 , A_1 , E и c (из которых, в основном, лишь первые три неизвестные представляют интерес):

$$\begin{cases} f(a) - A_0 - A_1 a = E, \\ f(c) - A_0 - A_1 c = -E, \\ f(b) - A_0 - A_1 b = E, \\ f'(c) - A_1 = 0 \end{cases} \quad \text{или} \quad \begin{cases} f(a) - A_0 - A_1 a = -E, \\ f(c) - A_0 - A_1 c = E, \\ f(b) - A_0 - A_1 b = -E, \\ f'(c) - A_1 = 0. \end{cases}$$

Получить решение такой системы в общем виде не представляется возможным, поскольку неизвестная величина c входит в нее нелинейным образом (в каждом конкретном случае подобная система без проблем решается численно).

Ключом к геометрической интерпретации случая Б служит двойное выражение коэффициента A_1 из уравнений системы: из первого и третьего уравнений, рассматриваемых совместно, имеем

$A_1 = \frac{f(b) - f(a)}{b - a}$, а из четвертого следует $A_1 = f'(c)$, что го-

ворит о параллельности хорды, стягивающей концы дуги графика $f(x)$ на $[a, b]$, касательной, проведенной к $f(x)$ во внутренней точке альтернанса, и аппроксимирующей $f(x)$ прямой $y = A_0 + A_1 x$ (рис. 2.4).

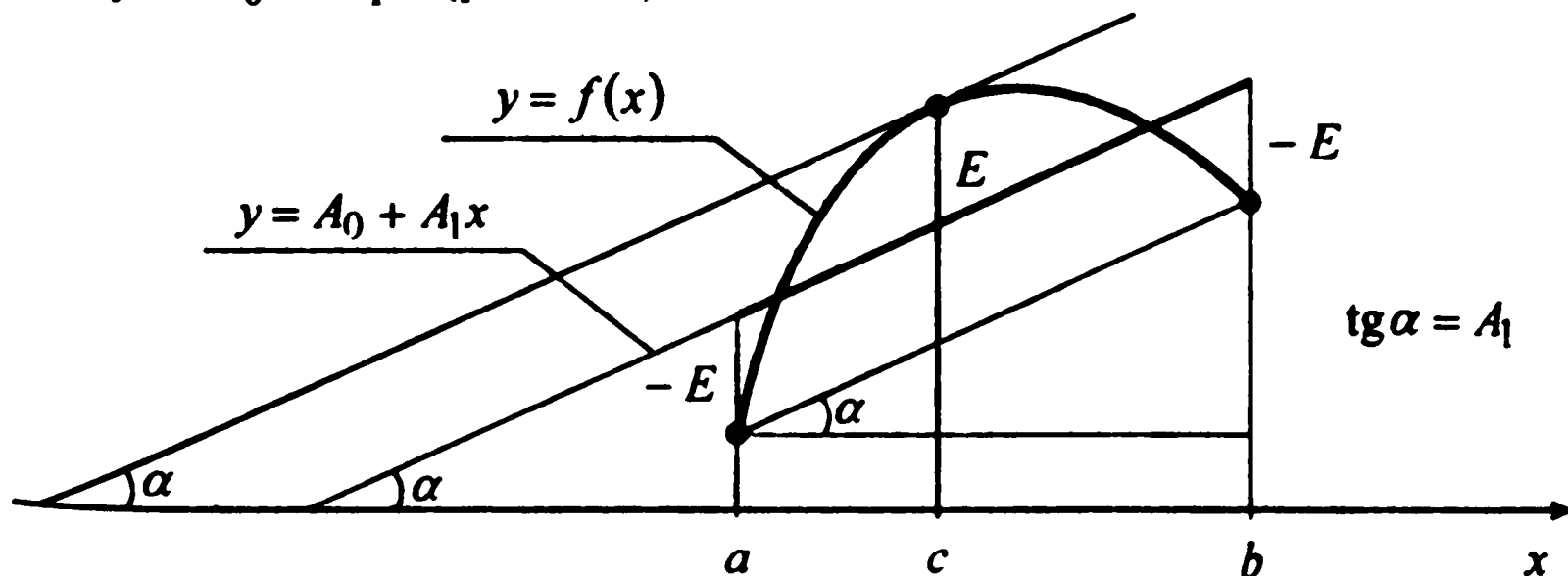


Рис. 2.4. *Наилучшая линейная аппроксимация дифференцируемой выпуклой функции*

Пример 2.1. Построим многочлены наилучшего равномерного приближения нулевой и первой степеней для функции $y = \sin x$ на отрезке $\left[0, \frac{\pi}{6}\right]$.

Сразу заметим, что данная функция всюду дифференцируема (а значит, непрерывна) и выпукла вверх на заданном отрезке. При этом

$$m := \min_{x \in [0, \frac{\pi}{6}]} \sin x = \sin 0 = 0, \quad M := \max_{x \in [0, \frac{\pi}{6}]} \sin x = \sin \frac{\pi}{6} = 0.5.$$

Следовательно, согласно рассмотренному выше случаю А, найдя

$$\frac{M + m}{2} = \frac{0.5 + 0}{2} = 0.25 \quad \text{и} \quad \frac{M - m}{2} = \frac{0.5 - 0}{2} = 0.25,$$

при $x \in \left[0, \frac{\pi}{6}\right]$ можно считать, что $\sin x \approx 0.25$ с предельной погрешностью 0.25.

Далее, в соответствии со случаем Б, продифференцировав данную функцию, составляем систему:

$$\begin{cases} \sin 0 - A_0 - A_1 \cdot 0 = -E, \\ \sin c - A_0 - A_1 c = E, \\ \sin \frac{\pi}{6} - A_0 - A_1 \frac{\pi}{6} = -E, \\ \cos c - A_1 = 0, \end{cases} \quad \text{т. е.} \quad \begin{cases} A_0 = E, \\ \sin c - A_0 - A_1 c = E, \\ 0.5 - A_0 - A_1 \frac{\pi}{6} = -E, \\ \cos c = A_1. \end{cases}$$

Из нее последовательно находим:

$$A_1 = 0.5 : \frac{\pi}{6} = \frac{3}{\pi} \approx 0.9549;$$

$$c = \arccos \frac{3}{\pi} \approx 0.3014;$$

$$A_0 (= E) = \frac{1}{2}(\sin c - A_1 c) \approx 0.0045.$$

Таким образом, функцию $y = \sin x$ на отрезке $\left[0, \frac{\pi}{6}\right]$ можно подменить линейной функцией $y = 0.0045 + 0.9549x$, и наибольшая ошибка при этом не будет превышать величины ≈ 0.0045 .

§ 2.4. ЭКОНОМИЗАЦИЯ СТЕПЕННЫХ РЯДОВ

Рассмотрим один из простейших способов построения многочленов, близких к наилучшим равномерным, о существовании которых упоминалось в предыдущем параграфе.

В математическом анализе хорошо изучено и широко применяется разложение функций в степенные ряды, в частности, в ряды Тейлора. Частичные суммы таких рядов — многочлены — используются в качестве локальных аппроксимаций для исходных функций. Степени используемых при этом многочленов зависят от требуемой точности аппроксимации, положения точки из области сходимости ряда, в окрестности которой производится аппроксимация, скорости сходимости ряда. В некоторых случаях такой подход мало приемлем из-за медленной сходимости рядов и большой неравномерности, т. е. существенной разницы в необходимых для заданной точности степенях приближающих многочленов при разных значениях аргумента.

Для улучшения указанных параметров частичных сумм степенных рядов можно привлечь многочлены Чебышева. Процедура преобразования степенного ряда, представляющего собой разложение некоторой функции по системе степенных функций, в разложение ее по системе многочленов Чебышева называется экономизацией степенного ряда [90].

Чтобы преобразовать степенной ряд в ряд по системе многочленов Чебышева, нужно сначала обратить формулы, по которым многочлены Чебышева $T_n(x)$ выражаются через степенные функции. А именно, глядя на несколько записанных в § 2.1 первых многочленов Чебышева, можно через них выразить степени x последовательно одну за другой:

$$\begin{aligned} 1 &= T_0; \\ x &= T_1; \\ x^2 &= \frac{1}{2}(T_0 + T_2); \\ x^3 &= \frac{1}{4}(3T_1 + T_3); \\ x^4 &= \frac{1}{8}(3T_0 + 4T_2 + T_4); \\ x^5 &= \frac{1}{16}(10T_1 + 5T_3 + T_5) \end{aligned} \tag{2.12}$$

и т. д. (Аргумент x в этих выражениях для краткости опущен).

Если некоторая функция $y = f(x)$ на некотором промежутке $[a, b] \subseteq [-1, 1]$ представлена степенным рядом

$$f(x) = a_0 + a_1x + a_2x^2 + \dots + a_kx^k + \dots,$$

то, подставляя сюда вместо степеней x их выражения (2.12) через многочлены Чебышева и приводя подобные члены, можно, в

принципе, получить разложение $f(x)$ вида

$$f(x) = b_0 + b_1 T_1 + b_2 T_2 + \dots + b_n T_n + \dots \quad (2.13)$$

Имея в виду рассмотренные в § 2.1 свойства многочленов Чебышева $T_n(x)$, можно рассчитывать, что многочлены, получаемые усечением разложений (2.13), будут близки к многочленам наилучших равномерных приближений.

Не пытаясь ответить на все возникающие здесь вопросы как теоретического, так и практического характера (некоторые сведения об этом можно найти, например, в [90]), посмотрим на примере, какой эффект может дать простейшая процедура экономизации.

Пример 2.2. Известно [28], что при $x \in (-1, 1]$ справедливо разложение

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \frac{x^5}{5} - \dots \quad (2.14)$$

Ограничиваясь первыми пятью членами этого разложения, т. е. многочленом пятой степени, перейдем с помощью равенств (2.12) к приближенному представлению функции $\ln(1+x)$ через многочлены Чебышева:

$$\begin{aligned} \ln(1+x) \approx & T_1 - \frac{1}{4}(T_0 + T_2) + \frac{1}{12}(3T_1 + T_3) - \frac{1}{32}(3T_0 + 4T_2 + T_4) + \\ & + \frac{1}{80}(10T_1 + 5T_3 + T_5) = -\frac{11}{32}T_0 + \frac{11}{8}T_1 - \frac{3}{8}T_2 + \frac{7}{48}T_3 - \frac{1}{32}T_4 + \frac{1}{80}T_5. \end{aligned}$$

Если здесь оставить всего три первых члена, т. е. представить $\ln(1+x)$ многочленом второй степени

$$\ln(1+x) \approx -\frac{11}{32}T_0 + \frac{11}{8}T_1 - \frac{3}{8}T_2 = \frac{1}{32} + \frac{11}{8}x - \frac{3}{4}x^2,$$

и сравнить полученное квадратичное приближение функции $\ln(1+x)$ с квадратичным же тейлоровским приближением $\ln(1+x) \approx x - \frac{x^2}{2}$, непосредственно вытекающим из разложения (2.14), то получим картину, изображенную на рис. 2.5.

Она показывает (как это и следовало ожидать), что в середине интервала сходимости ряда (2.14) данная функция точнее аппроксимируется функцией $y = x - \frac{x^2}{2}$, а вблизи концов этого интервала — функцией $y = \frac{1}{32} + \frac{11}{8}x - \frac{3}{4}x^2$. Максимальная на $(-1, 1]$ погрешность в последнем случае меньше, но, разумеется, полученная экономизацией квадратичная функция не является многочленом наилучшего равномерного приближения среди всех многочленов второй степени.

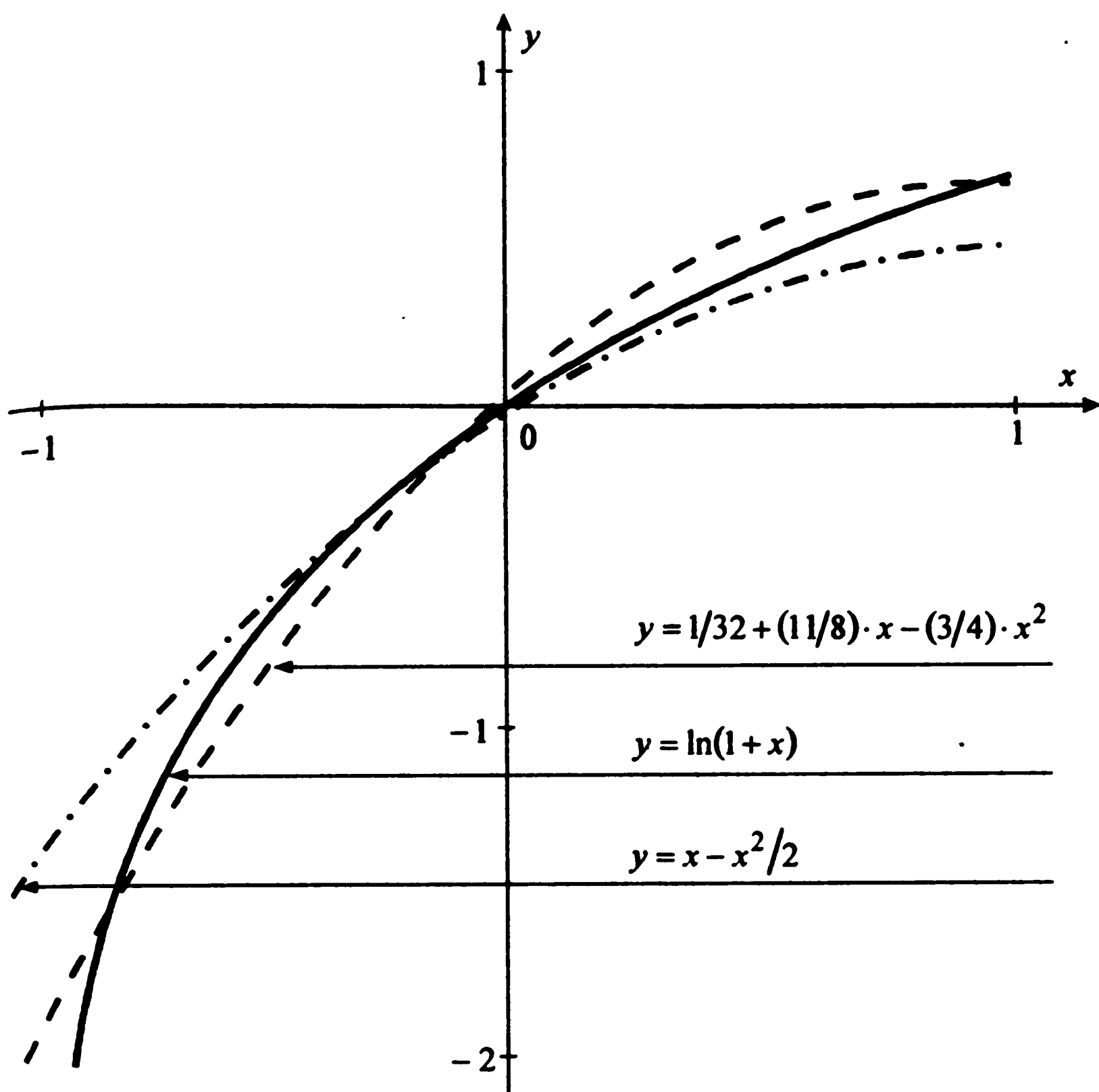


Рис. 2.5. Графики функции $\ln(1+x)$ и двух ее квадратичных приближений

Многочлены наилучших равномерных приближений вызывают большой теоретический и практический интерес. Например, вычисление значений некоторых основных элементарных функций на микрокалькуляторах и компьютерах базируется на том, что подсчет значения функции в произвольной точке области определения сводится к вычислению значения на некотором стандартном промежутке, на котором данная функция подменяется многочленом, близким к многочлену наилучшего равномерного приближения такой степени, при которой гарантируется, что максимальная ошибка не будет превосходить заданной фиксированной величины при любом значении аргумента из этого промежутка [30, 61].

Кроме процедуры экономизации степенных рядов, для построения многочленов, близких к многочленам наилучших равномерных приближений, привлекают и другие приемы, из которых наиболее известны алгоритмы Ремеза и Валле–Пуссена [32, 71, 75, 78].¹

Замечание 2.2. Можно встретить и несколько иное трактование понятия (и процедуры) *экономизации степенных разложений*. Так, например, в [13] (см. также [21]) этим термином определяется процесс перехода от аппроксимации функции $f(x)$ многочленом n -й степени к аппроксимации многочленом $(n-1)$ -й степени с сохранением (оценки) точности равномерного приближения. В основе такого процесса лежит следующее утверждение.

Теорема 2.3. Пусть $f(x) \approx S_n(x) := \sum_{k=0}^n a_k x^k$, причем

$$\max_{x \in [-1, 1]} |f(x) - S_n(x)| \leq \delta < \varepsilon.$$

Тогда, если выполняется неравенство $\delta + \frac{|a_n|}{2^{n-1}} < \varepsilon$, то многочлен

$P_{n-1}(x) := S_{n-1}(x) + a_n \left(x^n - \frac{1}{2^{n-1}} T_n(x) \right)$ аппроксимирует $f(x)$ на отрезке $[-1, 1]$ с оценкой

$$\max_{x \in [-1, 1]} |f(x) - P_{n-1}(x)| < \varepsilon.$$

УПРАЖНЕНИЯ

2.1. Запишите несколько первых смещенных многочленов Чебышева (см. замечание 2.1). Какова формула корней смещенного многочлена Чебышева n -й степени?

2.2. Пусть функция $f(x)$ определена и достаточное число раз дифференцируема на отрезке $[2, 4]$. В каких точках следует знать значения $f(x)$, чтобы проинтерполировать ее с минимальной максимальной погрешностью:

- а) многочленом первой степени?
- б) многочленом второй степени?

2.3. Найдите неулучшаемую оценку погрешности интерполирования функции $f(x) = \ln(1+x)$ на отрезке $[0, 1]$:

- а) многочленом первой степени;
- б) многочленом второй степени.

Постройте эти многочлены.

2.4. Постройте многочлены наилучшего равномерного приближения нулевой и первой степеней и укажите наибольшие величины погрешностей аппроксимации для функций:

а) $f(x) = \sqrt{x}$ при $x \in [0, 1]$;

б) $f(x) = \frac{1}{x}$ при $x \in [1, 2]$;

в) $f(x) = \ln(1 + x)$ при $x \in [0, 1]$;

В случае в) сравните чебышевское приближение первой степени с результатом линейной интерполяции по чебышевским узлам (см. упр. 2.3).

2.5. Используя разложение функции $f(x) = \operatorname{arctg} x$ в ряд Тейлора (Маклорена) до пятой степени x , примените процедуру экономизации для представления этой функции через многочлены Чебышева первой и третьей степеней. Дайте графическое сравнение результатов такой экономизации, аналогичное приведенному на рис. 2.5.

ГЛАВА 3 || МЕТОД НАИМЕНЬШИХ || КВАДРАТОВ И НАИЛУЧШИЕ || СРЕДНЕКВАДРАТИЧЕСКИЕ || ПРИБЛИЖЕНИЯ

Рассматриваются две простейшие задачи, приводящие к широко известному методу наименьших квадратов: решение неопределенных линейных алгебраических систем и подбор параметров эмпирических зависимостей. Далее ставится более общая задача аппроксимации функций обобщенными многочленами, минимизирующими среднеквадратические отклонения в той или иной метрике; показывается, с какими проблемами при этом можно столкнуться, если выполнять полиномиальную аппроксимацию на базе обычных степенных функций. Приводятся справочные данные о классических ортогональных многочленах, изучается простая схема построения системы ортогональных многочленов, выписываются конкретные формулы для получения с их помощью многочленов Фурье, аппроксимирующих заданные функции в интегральном или в точечном смысле.

§ 3.1. ПРОСТЕЙШАЯ ОБРАБОТКА ЭМПИРИЧЕСКИХ ДАННЫХ МЕТОДОМ НАИМЕНЬШИХ КВАДРАТОВ

Рассмотрим две задачи, возникающие вне математики при изучении каких-либо закономерностей на основе наблюдений или экспериментов и приводящие к истокам широко известного и весьма популярного метода наименьших квадратов^{*)}. Характерной особенностью таких задач является то обстоятельство, что исходный материал для выявления тех или иных закономерностей — заведомо приближенный, в силу неточности измерительных приборов, неповторимости условий наблюдений, случайных ошибок и пр. Отсюда естественность стремления исследователей к накоплению как можно большего количества экспериментальных данных с тем, чтобы, усредняя их каким-либо образом, получить требуемые значения, как правило, небольшого числа параметров изучаемых закономерностей наиболее точно.

^{*)} Разработку этого метода связывают с именами знаменитых математиков прошлого К. Гаусса (1777–1855) и А. Лежандра (1752–1833).

Пример 3.1. Пусть определению подлежат две величины x_1 и x_2 , связь между которыми задается тремя равенствами:

$$\begin{cases} x_1 + 2x_2 = 8.5, \\ 2x_1 + x_2 = 6.7, \\ x_1 + 3x_2 = 11.1, \end{cases} \quad (3.3)$$

и пусть правые части уравнений (3.3) — заведомо приближенные числа. Совершенно очевидно, что эта система противоречива, т.е. не имеет решения в обычном смысле. Для получения ее нормального псевдорешения, т.е. решения в смысле метода наименьших квадратов, составим новую систему по типу (3.2) при $m = 2$, $n = 3$:

$$\begin{cases} 6x_1 + 7x_2 = 33, \\ 7x_1 + 14x_2 = 57. \end{cases}$$

Решив последнюю, получаем пару $x_1^* = 1.8$, $x_2^* \approx 3.171$, не удовлетворяющую никакому отдельно взятому уравнению данной системы (3.3) и, в то же время, устраивающую систему в целом в оговоренном смысле. Геометрическое представление о найденном решении можно получить из рис. 3.1.

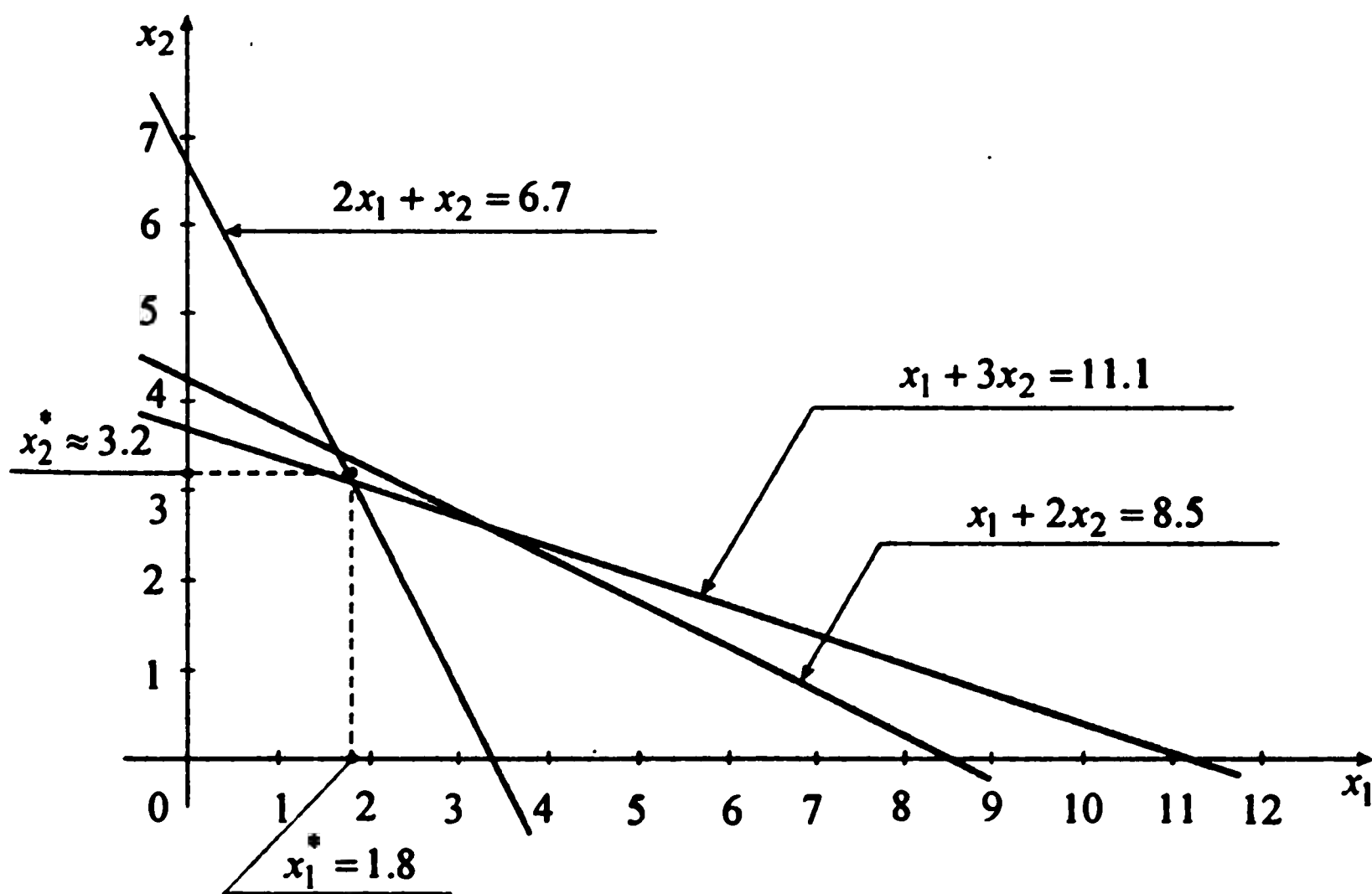


Рис. 3.1. Геометрическая интерпретация системы (3.3) и ее нормального псевдорешения

Замечание 3.1. Геометрические образы уравнений системы (3.3) не изменяются при умножении левых и правых частей уравнений на одинаковые отличные от нуля числа, т.е. можно считать, что преобразованная таким способом система в этом смысле эквивалентна исходной. Однако находимые методом наименьших квадратов нормальные псевдорешения эквивалентных в указанном смысле систем будут различаться. Например,

если второе уравнение системы (3.3) умножить на 100, т.е. заменить уравнением

$$200x_1 + 100x_2 = 670,$$

а первое и третье оставить неизменными, то система вида (3.2) для такого случая есть

$$\begin{cases} 40002x_1 + 20005x_2 = 134019.6, \\ 20005x_1 + 10013x_2 = 67050.3. \end{cases}$$

Решая ее, получаем псевдорешение $\tilde{x}_1 \approx 1.756$, $\tilde{x}_2 \approx 3.188$, близкое к найденному выше, но все же отличное от него.

Невязки r_i двумерных линейных уравнений можно интерпретировать как отклонения точки от прямых в тех случаях, когда их уравнения приводятся к нормальному виду. Деля каждое уравнение данной системы на квадратный корень из суммы квадратов его коэффициентов, приходим сначала к нормализованной системе

$$\begin{cases} \frac{1}{\sqrt{5}}x_1 + \frac{2}{\sqrt{5}}x_2 = \frac{8.5}{\sqrt{5}}, \\ \frac{2}{\sqrt{5}}x_1 + \frac{1}{\sqrt{5}}x_2 = \frac{6.7}{\sqrt{5}}, \\ \frac{1}{\sqrt{10}}x_1 + \frac{3}{\sqrt{10}}x_2 = \frac{11.1}{\sqrt{10}}, \end{cases}$$

из нее, как и выше, получаем однозначно разрешимую систему

$$\begin{cases} 1.1x_1 + 1.1x_2 = 5.49, \\ 1.1x_1 + 1.9x_2 = 8.07, \end{cases}$$

решение которой $\bar{x}_1 \approx 1.766$, $\bar{x}_2 \approx 3.225$ имеет хорошую геометрическую интерпретацию: сумма квадратов расстояний от точки $(\bar{x}_1; \bar{x}_2)$ до всех прямых, определяемых уравнениями данной системы (3.3), меньше, чем от любой другой точки $(x_1; x_2)$.

Ответ на вопрос о том, стоит ли искать такое оптимальное псевдорешение с помощью нормализации уравнений данной системы, или составлять систему вида (3.2) непосредственно из данной системы, или проводить перед ее составлением некоторое предварительное масштабирование (усиливающее или уменьшающее роль отдельных связей между искомыми величинами), в конкретных случаях зависит от содержательного смысла коэффициентов и неизвестных СЛАУ (3.1).

Задача Б. Предположим, что между независимой переменной x и зависимой переменной y имеется некая неизвестная функциональная связь $y = f(x)$. Эта связь отображается таблицей

x	x_0	x_1	$\dots$	x_n
y	y_0	y_1	$\dots$	y_n

приближенных значений $y_i \approx f(x_i)$, получаемых в ходе наблюдений или экспериментов. Требуется дать приближенное аналитическое описание этой связи, т.е. подобрать функцию $\varphi(x)$

представляющей необходимые условия экстремума функции $\Phi(a_1, a_2, \dots, a_m)$, в силу ее специфики, являющиеся и достаточными условиями ее минимума.

Если функция $\varphi(x, a_1, a_2, \dots, a_m)$ есть линейная функция относительно своих параметров $a_1, a_2, \dots, a_m$, то система (3.4) тоже будет линейной; в общем случае (3.4) — нелинейная система, что влечет за собой определенные трудности при ее решении. Спасительным в последней ситуации является тот факт, что обычно при задании семейств функций, аппроксимирующих реальные зависимости, число параметров берется небольшим (2 – 3), причем какие-то из этих параметров могут входить линейным образом.

В зависимости от характера табличных данных, изучаемого с помощью их изображения в соответствующей системе координат или с помощью некоторых прикидочных расчетов (см., например, [27, 31]), при обработке результатов экспериментов часто используют те или иные из следующих двухпараметрических семейств функций:

$$\begin{aligned} y &= ax + b, & y &= a + b \ln x & (y &= a + b \lg x), \\ y &= ax^b, & y &= ae^{bx} & (y &= a \cdot 10^{bx}), \\ y &= a + \frac{b}{x}, & y &= \frac{1}{ax + b}, & y &= \frac{x}{ax + b}; \end{aligned}$$

реже применяются трехпараметрические семейства

$$\begin{aligned} y &= ax^2 + bx + c, & y &= ax^b + c, \\ y &= ae^{bx} + c & (y &= a \cdot 10^{bx} + c); \end{aligned}$$

при изучении периодических явлений применяют тригонометрические функции.

В качестве примера рассмотрим ситуацию, когда есть основания считать, что значения x_i и соответствующие им приближенные значения y_i отражают некую линейную зависимость. Тогда, полагая

$$\varphi(x, a, b) = ax + b,$$

находим $\frac{\partial \varphi}{\partial a} = x$, $\frac{\partial \varphi}{\partial b} = 1$, и для вычисления параметров этой функции φ составляем систему типа (3.4):

$$\begin{cases} \sum_{i=0}^n (ax_i + b - y_i) \cdot x_i = 0, \\ \sum_{i=0}^n (ax_i + b - y_i) \cdot 1 = 0. \end{cases}$$

упрощая ее, приходим к стандартной СЛАУ

$$\begin{cases} \left(\sum_{i=0}^n x_i^2 \right) a + \left(\sum_{i=0}^n x_i \right) b = \sum_{i=0}^n x_i y_i, \\ \left(\sum_{i=0}^n x_i \right) a + (n+1)b = \sum_{i=0}^n y_i \end{cases} \quad (3.5)$$

с симметричной квадратной матрицей коэффициентов и заведомо однозначным решением $(a^*; b^*)$.

Заметим, что вместо того, чтобы решать нелинейные системы, получающиеся из (3.4) при поиске параметров конкретных семейств функций, когда эти параметры входят туда нелинейным образом, можно попытаться сначала линеаризовать подбираемую зависимость.

Пусть, например, в качестве аппроксимирующей $f(x)$ функции берется экспоненциальная функция

$$\varphi(x, a, b) = ae^{bx}. \quad (3.6)$$

Найдя ее частные производные

$$\frac{\partial \varphi}{\partial a} = e^{bx}, \quad \frac{\partial \varphi}{\partial b} = a b e^{bx},$$

видим, что подстановка этой функции и ее производных по параметрам в (3.4) приведет к нелинейной относительно параметра b системе. Чтобы не решать нелинейных уравнений, поступим так.

Прологарифмируем равенство

$$ae^{bx} = y,$$

получив при этом

$$\ln a + bx = \ln y,$$

введем новый параметр $A := \ln a$ и пересчитаем данную таблицу, переводя значения y_i в $Y_i := \ln y_i$. По таблице

x	x_0	x_1	$\dots$	x_n
Y	Y_0	Y_1	$\dots$	Y_n

методом наименьших квадратов находим оптимальные параметры A^* и b^* линейной функции $Y = A + bx$ (полагая $\varphi_1(x, A, b) = A + bx$, пользуемся системой (3.5), считая там $a = b$, а $b = A$). После этого вычисляем оптимальное значение $a^* = e^{A^*}$ параметра a исходной зависимости (3.6) и записываем итоговый результат

$$f(x) \approx a^* e^{b^* x}.$$

Замечание 3.2. К методу наименьших квадратов могут привести и иные соображения. Например, в курсах математической статистики к нему приходят, изучая корреляционные и функциональные зависимости при наличии случайных ошибок [15, 38, 44 и др.].

§ 3.2. ОБОБЩЕННЫЕ МНОГОЧЛЕНЫ НАИЛУЧШИХ СРЕДНЕКВАДРАТИЧЕСКИХ ПРИБЛИЖЕНИЙ

Вернемся к общей постановке задачи аппроксимации функций.

Пусть аппроксимируемая функция $f(x)$ и аппроксимирующая функция $\varphi(x)$ непрерывны на отрезке $[a, b]$ и аппроксимация должна производиться так, чтобы функция $\varphi(x)$ «в среднем хорошо описывала» поведение функции $f(x)$ при $x \in [a, b]$. Будем здесь постоянно иметь в виду две аппроксимационные ситуации: первая, это когда функция $f(x)$ считается (по крайней мере, теоретически) известной в любой точке x отрезка $[a, b]$, и близость между $f(x)$ и $\varphi(x)$ понимается в интегральном смысле, и вторая, когда $f(x)$ известна (причем приближенно) только в $n+1$ точках $x_0, x_1, \dots, x_n$ отрезка $[a, b]$, в которых и производится согласование $f(x)$ с $\varphi(x)$ подобно тому, как это делалось при рассмотрении задачи Б в предыдущем параграфе. В связи с этим, будем параллельно рассматривать:

а) пространство $C_L[a, b]$ непрерывных на $[a, b]$ функций со скалярным произведением

$$(f, \varphi)_{C_L[a, b]} := \frac{1}{b-a} \int_a^b f(x) \varphi(x) dx, \quad (3.7)$$

метрикой (расстоянием)

$$\rho(f, \varphi)_{C_L[a, b]} := \sqrt{\frac{1}{b-a} \int_a^b (f(x) - \varphi(x))^2 dx} \quad (3.8)$$

и нормой

$$\|f\|_{C_L[a, b]} := \rho(f, 0)_{C_L[a, b]} = \sqrt{(f, f)_{C_L[a, b]}} = \sqrt{\frac{1}{b-a} \int_a^b f^2(x) dx};$$

б) пространство $R_{n+1}[a, b]$ сеточных функций, определенных в точках $x_i \in [a, b]$ ($i = 0, 1, \dots, n$), со скалярным произведением

$$(f, \varphi)_{R_{n+1}[a, b]} := \frac{1}{n+1} \sum_{i=0}^n f(x_i) \varphi(x_i), \quad (3.9)$$

метрикой

$$\rho(f, \varphi)_{R_{n+1}[a, b]} := \sqrt{\frac{1}{n+1} \sum_{i=0}^n (f(x_i) - \varphi(x_i))^2} \quad (3.10)$$

и нормой

$$\|f\|_{R_{n+1}[a, b]} := \rho(f, 0)_{R_{n+1}[a, b]} = \sqrt{(f, f)_{R_{n+1}[a, b]}} = \sqrt{\frac{1}{n+1} \sum_{i=0}^n f^2(x_i)}.$$

Введенные указанным способом метрики (3.8) и (3.10) характеризуют близость функций $f(x)$ и $\varphi(x)$ в пространствах $C_L[a, b]$ и $R_{n+1}[a, b]$ соответственно; по отношению к приближенному равенству $f(x) \approx \varphi(x)$ при $x \in [a, b]$ они представляют собой **интегральную** и **точечную** (дискретную) **среднеквадратические ошибки**.

В силу того, что

$$\sum_{i=0}^n (f(x_i) - \varphi(x_i))^2 = \min \Leftrightarrow \rho(f, \varphi)_{R_{n+1}[a, b]} = \min$$

и

$$\int_a^b (f(x) - \varphi(x))^2 dx = \min \Leftrightarrow \rho(f, \varphi)_{C_L[a, b]} = \min,$$

можно сказать, что применение метода наименьших квадратов к аппроксимации функции $f(x)$ функцией $\varphi(x)$ заданного семейства означает подбор такой функции $\varphi(x)$, которая минимизирует среднеквадратическую погрешность приближенного равенства $f(x) \approx \varphi(x)$ в интегральном или в точечном смысле; в связи с этим, она называется **наилучшим среднеквадратическим приближением** $f(x)$ на заданном семействе функций.

Рассмотрим, к чему сводится процесс построения наилучших среднеквадратических приближений в одном конкретном, но достаточно общем случае, когда аппроксимирующая $f(x)$ функция $\varphi(x)$ представляет собой линейную комбинацию нескольких других, вообще говоря, более простых (базисных) функций.

Пусть $\{\varphi_j(x)\}_{j=0}^m$ — некоторая заданная на $[a, b]$ система линейно независимых функций. **Обобщенным многочленом** будем называть функцию

$$Q_m(x) := c_0 \varphi_0(x) + c_1 \varphi_1(x) + \dots + c_m \varphi_m(x), \quad (3.11)$$

где $c_0, c_1, \dots, c_m$ — произвольные вещественные числа (коэффициенты обобщенного многочлена). Поскольку функции $\varphi_j(x)$ считаются заданными, построение **обобщенного многочлена наилучшего среднеквадратического приближения** для данной функции $f(x)$ сводится к нахождению оптимального набора $c_0^*, c_1^*, \dots, c_m^*$ коэффициентов $Q_m(x)$ в (3.11) на основе метода

наименьших квадратов, т.е. к решению задач минимизации:

$$\sum_{i=0}^n (c_0 \varphi_0(x_i) + c_1 \varphi_1(x_i) + \dots + c_m \varphi_m(x_i) - f(x_i))^2 \rightarrow \min \quad (3.12)$$

— в дискретном случае — и

$$\int_a^b (c_0 \varphi_0(x) + c_1 \varphi_1(x) + \dots + c_m \varphi_m(x) - f(x))^2 dx \rightarrow \min \quad (3.13)$$

— в непрерывном случае.

Для первой из этих задач, т.е. для (3.12), необходимые (и достаточные) условия выражаются системой

$$\left\{ \begin{array}{l} \sum_{i=0}^n \varphi_0(x_i)(c_0\varphi_0(x_i) + c_1\varphi_1(x_i) + \dots + c_m\varphi_m(x_i) - f(x_i)) = 0, \\ \sum_{i=0}^n \varphi_1(x_i)(c_0\varphi_0(x_i) + c_1\varphi_1(x_i) + \dots + c_m\varphi_m(x_i) - f(x_i)) = 0, \\ \dots\dots\dots \\ \sum_{i=0}^n \varphi_m(x_i)(c_0\varphi_0(x_i) + c_1\varphi_1(x_i) + \dots + c_m\varphi_m(x_i) - f(x_i)) = 0. \end{array} \right. \quad (3.14)$$

После элементарных преобразований она может быть переписана в терминах скалярных произведений (см. (3.10)):

[illegible]

Если сеточные функции $\varphi_j(x_i)$ образуют систему линейно независимых элементов пространства $R_{n+1}[a, b]$, то полученная симметричная линейная алгебраическая система (3.15), называемая **нормальной системой МНК**, имеет заведомо отличный от нуля определитель (это известный определитель Грама [2, 9, 26]), и значит, однозначно разрешима. Следовательно, при заданном базисе $\{\varphi_j\}$ путем решения системы (3.15) можно найти единственный обобщенный многочлен

$$Q_m^*(x) := c_0^* \varphi_0(x) + c_1^* \varphi_1(x) + \dots + c_m^* \varphi_m(x)$$

такой, что $f(x) \approx Q_m^*(x)$ при $x \in [a, b]$ с наименьшей среднеквадратической ошибкой

$$\rho(f, Q_m^*(x)) = \sqrt{\frac{1}{n+1} \sum_{i=0}^n (f(x_i) - Q_m^*(x_i))^2}. \quad (3.16)$$

Вернемся вновь на начальный этап построения обобщенного многочлена наилучшего среднеквадратического приближения, но теперь уже в интегральном смысле. Легко убедиться, что записав для задачи (3.13) необходимые условия, т.е. получив интегральный аналог системы (3.14), от нее приходим к той же самой нормальной системе МНК (3.15), только скалярные произведения здесь расшифровываются с помощью формулы (3.7). Для той же $f(x)$ и при том же базисе $\{\varphi_j\}$, ввиду различий в скалярных произведениях (3.7) и (3.9), решение системы (3.15) в этом случае дает, вообще говоря, уже другой (возможно, «близкий» к $c_0^*, c_1^*, \dots, c_m^*$) оптимальный набор $\tilde{c}_0, \tilde{c}_1, \dots, \tilde{c}_m$ коэффициентов (3.11) и приводит к представлению

$$f(x) \approx \tilde{Q}_m(x) = \tilde{c}_0 \varphi_0(x) + \tilde{c}_1 \varphi_1(x) + \dots + \tilde{c}_m \varphi_m(x)$$

с наименьшим среднеквадратическим отклонением

$$\rho(f, \tilde{Q}_m(x)) = \sqrt{\frac{1}{b-a} \int_a^b (f(x) - \tilde{Q}_m(x))^2 dx} \quad (3.17)$$

(т.е. гарантируется, что $\rho(f, \tilde{Q}_m(x))_{C_L[a,b]} \leq \rho(f, Q_m(x))_{C_L[a,b]}$, где $Q_m(x)$ — произвольная функция вида (3.11)).

Замечание 3.3. Множители $\frac{1}{b-a}$ в скалярном произведении (3.7) и

$\frac{1}{n+1}$ в (3.9) введены только ради удобства трактовки величин $\rho(f, \varphi)$ в (3.8) и в (3.10) как среднеквадратических ошибок приближенного равенства $f(x) \approx \varphi(x)$. Пользуясь более привычными формулами скалярных произведений (без этих множителей), в итоге все равно пришли бы к той же системе (3.15) (проверьте!). Так что в дальнейшем под скалярными произведениями в нормальных системах МНК могут пониматься обычные интегральные и точечные (евклидовы) скалярные произведения, в то время как при подсчете среднеквадратических погрешностей (т.е. в формулах (3.16), (3.17)) эти множители обязаны присутствовать.

Анализируя СЛАУ (3.15), приходим к выводу, что она чрезвычайно упрощается в случае, когда базисные функции $\varphi_j(x)$ образуют на $[a, b]$ ортогональную систему.

Как известно, взаимная ортогональность функций из множества $\{\varphi_j(x)\}_{j=0}^m$ означает, что $(\varphi_j, \varphi_k) = 0$ при любых $k \neq j$. Следовательно, коэффициенты $c_0, c_1, \dots, c_m$ обобщенного многочлена наилучшего среднеквадратического приближения (3.11) могут быть сразу выписаны из превратившейся в диагональную

системы (3.15), а именно:

$$\begin{cases} c_0 = \frac{(\varphi_0, f)}{(\varphi_0, \varphi_0)} = \frac{(\varphi_0, f)}{\|\varphi_0\|^2}, \\ c_1 = \frac{(\varphi_1, f)}{(\varphi_1, \varphi_1)} = \frac{(\varphi_1, f)}{\|\varphi_1\|^2}, \\ \dots\dots\dots \\ c_m = \frac{(\varphi_m, f)}{(\varphi_m, \varphi_m)} = \frac{(\varphi_m, f)}{\|\varphi_m\|^2}. \end{cases} \quad (3.18)$$

В таком случае эти оптимальные коэффициенты называются **коэффициентами Фурье**, а обобщенный многочлен (3.11) с этим набором коэффициентов (т.е. обобщенный многочлен наилучшего среднеквадратического приближения для $f(x)$) называется **обобщенным многочленом Фурье**.

Еще проще построение таких многочленов, когда система $\{\varphi_j(x)\}_{j=0}^m$ — ортонормированная. Тогда $\|\varphi_j\|=1$, и из (3.18) следует, что $c_j = (\varphi_j, f)$ при любом $j \in \{0, 1, \dots, m\}$, т.е. аппроксимация функции $f(x)$ обобщенным многочленом Фурье имеет вид

$$f(x) \approx (\varphi_0, f)\varphi_0(x) + (\varphi_1, f)\varphi_1(x) + \dots + (\varphi_m, f)\varphi_m(x). \quad (3.19)$$

В частности, в математическом анализе достаточно подробно изучаются представления функций (не обязательно непрерывных) выражениями типа (3.19), в которых в качестве базисных функций используются функции ортогональной на $[-\pi, \pi]$ системы $\{\sin kx, \cos kx \mid k \in \mathbb{N}_0\}$. Такие представления в этом случае называются **отрезками тригонометрических рядов Фурье**.

§ 3.3. О НОРМАЛЬНОЙ СИСТЕМЕ МНК ПРИ ПОЛИНОМИАЛЬНОЙ АППРОКСИМАЦИИ

Возьмем в качестве базисных функций для обобщенного многочлена (3.11) степенные функции

$$\varphi_0 = 1, \quad \varphi_1 = x, \quad \varphi_2 = x^2, \quad \dots, \quad \varphi_m = x^m.$$

В таком случае он превращается в обычный многочлен степени m канонического вида:

$$Q_m(x) \equiv P_m(x) := c_0 + c_1x + c_2x^2 + \dots + c_mx^m. \quad (3.20)$$

число точек согласования $(n+1)$ было не меньшим числа отыскиваемых параметров $(m+1)$. При этом очевидно, что случай $m=n$ замыкает теорию полиномиальной аппроксимации по методу наименьших квадратов на теорию лагранжевой интерполяции.

§ 3.4. СИСТЕМЫ ОРТОГОНАЛЬНЫХ МНОГОЧЛЕНОВ

Имея в виду, что нас интересует, в первую очередь, полиномиальная аппроксимация функций, и при этом показаны недостатки степенных функций для их использования в качестве базисных при построении многочленов наилучших среднеквадратических приближений (§ 3.3) и отмечены достоинства ортогональных многочленов в этой роли (§ 3.2), приведем некоторые сведения справочного характера о последних.

Многочлены Лежандра $\chi_n(x)$ — наиболее употребительные из классических ортогональных многочленов и единственные, для которых условие их ортогональности на отрезке $[-1, 1]$ выполняется «в чистом виде», т.е. через равенство нулю скалярного произведения вида

$$(f, \varphi) := \int_{-1}^1 f(x)\varphi(x)dx, \quad (3.23)$$

а именно:

$$\int_{-1}^1 \chi_k(x)\chi_j(x)dx = \begin{cases} 0, & \text{если } k \neq j, \\ \frac{2}{2k+1}, & \text{если } k = j. \end{cases} \quad (3.24)$$

Для многочленов Лежандра известна явная, порождающая их, формула Родрига^{*)}

$$\chi_n(x) := \frac{1}{n! 2^n} \cdot \frac{d^n}{dx^n} (x^2 - 1)^n. \quad (3.25)$$

Положив $\chi_0 = 1$ и найдя по формуле (3.25) $\chi_1 = x$, все последующие многочлены Лежандра можно получать один за другим с помощью рекуррентной формулы

$$(n+1)\chi_{n+1}(x) - (2n+1)x\chi_n(x) + n\chi_{n-1}(x) = 0. \quad (3.26)$$

^{*)} Родриг (Родригес) Бенжамен Оленд (1794–1851) — французский математик. Появление формулы Родрига (3.25) датируется 1814 годом.

При $n = 1, 2, 3, \dots$ из нее имеем соответственно:

$$\chi_2 = \frac{1}{2}(3x^2 - 1),$$

$$\chi_3 = \frac{1}{2}(5x^3 - 3x),$$

$$\chi_4 = \frac{1}{8}(35x^4 - 30x^2 + 3),$$

$$\chi_5 = \frac{1}{8}(63x^5 - 70x^3 + 15x) \quad \text{и т. д.}$$

Сравнивая многочлены Лежандра с рассмотренными в гл. 2 многочленами Чебышева, можно обнаружить, что их сходство — не только внешнее. Эти многочлены характеризует ряд одинаковых свойств. Например, как и многочлены Чебышева, многочлены Лежандра n -й степени имеют на отрезке $[-1, 1]$ ровно n различных действительных корней (см. свойство 2.1). В то же время, как и многочлены Лежандра, многочлены Чебышева $T_n(x)$ относятся к числу классических ортогональных многочленов, но ортогональность здесь понимается в более широком смысле — это **ортогональность с весом**. А именно, условие ортогональности многочленов Чебышева $T_n(x)$ на отрезке $[-1, 1]$ имеет вид

$$\int_{-1}^1 T_k(x) \cdot T_j(x) \frac{dx}{\sqrt{1-x^2}} = \begin{cases} 0, & \text{если } k \neq j, \\ \frac{\pi}{2}, & \text{если } k = j \neq 0, \\ \pi, & \text{если } k = j = 0. \end{cases}$$

Наряду с многочленами $T_n(x)$, называемыми также **многочленами Чебышева первого рода**, ортогональными на $(-1, 1)$ являются и **многочлены Чебышева второго рода** $U_n(x)$, определяемые через $T_n(x)$, а именно:

$$U_n(x) := \frac{1}{n+1} T'_{n+1}(x) = \frac{\sin[(n+1)\arccos x]}{\sin \arccos x}.$$

Многочлены Лежандра и многочлены Чебышева первого и второго рода принадлежат одному семейству **многочленов Якоби**, ортогональных на конечном промежутке (a, b) с некоторой весовой функцией $p(x) > 0$ определенного вида (для многочленов Лежандра $p(x) \equiv 1$) и задаваемых формулой, обобщающей формулу Родрига (3.25) [69 и др.].

Многочлены Лагерра^{*)} $L_n(x)$ определим требованием их ортогональности на промежутке $[0, +\infty)$ с весовой функцией e^{-x} , а именно, условием

$$\int_0^{+\infty} e^{-x} L_k(x) \cdot L_j(x) dx = \begin{cases} 0, & \text{если } k \neq j, \\ (k!)^2, & \text{если } k = j. \end{cases}$$

Как и предыдущие ортогональные многочлены, многочлены Лагерра удовлетворяют рекуррентному соотношению типа (3.26)

$$L_{n+1}(x) - (2n+1-x)L_n(x) + n^2 L_{n-1}(x) = 0, \quad (3.27)$$

где $n \in N$, $L_0 := 1$, $L_1 := -x+1$. Отсюда легко получить несколько первых многочленов Лагерра:

$$L_2 = x^2 - 4x + 2,$$

$$L_3 = -x^3 + 9x^2 - 18x + 6,$$

$$L_4 = x^4 - 16x^3 + 72x^2 - 96x + 24,$$

$$L_5 = -x^5 + 25x^4 - 200x^3 + 600x^2 - 600x + 120 \quad \text{и т. д.}$$

Заметим, что под тем же обозначением $L_n(x)$ в литературе может встретиться многочлен $\frac{1}{n!}L_n(x)$. Кроме того, следует упомянуть, что существуют ортогональные многочлены Лагерра несколько более общего вида; в частности, ортогональность может пониматься с весовой функцией $p(x) = x^\alpha e^{-x}$, где $\alpha > -1$ (в рассмотренном случае $\alpha = 0$).

Многочлены Эрмита^{**)} $H_n(x)$ ортогональны на всей числовой оси с весом $p(x) = e^{-x^2}$. Они удовлетворяют интегральному условию ортогональности

$$\int_{-\infty}^{+\infty} e^{-x^2} H_k(x) \cdot H_j(x) dx = \begin{cases} 0, & \text{если } k \neq j, \\ 2^k k! \sqrt{\pi}, & \text{если } k = j, \end{cases}$$

и рекуррентному соотношению

$$H_{n+1}(x) - 2xH_n(x) + 2nH_{n-1}(x) = 0, \quad (3.28)$$

^{*)} Лагέρр Эдмон Никола (1834–1886) — французский математик.

^{**)} Не надо путать определяемые здесь ортогональные многочлены Эрмита с рассмотренными ранее интерполяционными многочленами Эрмита (см. § 1.8). Обозначение $H_n(x)$ (как и $L_n(x)$) несет в этой книге двойную нагрузку.

где $n \in \mathbb{N}$, $H_0 := 1$, $H_1 := 2x$. Подставляя в (3.28) последовательно $n = 1, 2, \dots$, находим первые многочлены Эрмита:

$$H_2 = 4x^2 - 2,$$

$$H_3 = 8x^3 - 12x,$$

$$H_4 = 16x^4 - 48x^2 + 12,$$

$$H_5 = 32x^5 - 160x^3 + 120x \quad \text{и т. д.}$$

Как видим, они имеют такую же структуру, как и многочлены Чебышева.

Ортогональные многочлены Якоби, Лагерра и Эрмита обладают рядом общих свойств. Все они являются решениями одного семейства обыкновенных дифференциальных уравнений второго порядка. Все ортогональные многочлены $P_n(x)$ одной и той же системы линейно независимы и удовлетворяют трехчленному рекуррентному соотношению вида

$$\alpha_n P_{n+1}(x) + (\beta_n - x)P_n(x) + \gamma_n P_{n-1}(x) = 0, \quad (3.29)$$

содержащему в себе, как легко убедиться, соотношения (3.26), (3.27), (3.28), а также (2.2), для конкретных многочленов Лежандра, Лагерра, Эрмита и Чебышева соответственно. Все n корней ортогональных многочленов $P_n(x)$ — простые и находятся на промежутке ортогональности; при этом корни многочлена $P_n(x)$ разделяются корнями многочлена $P_{n-1}(x)$ [90].

Более подробные сведения об ортогональных многочленах и их свойствах можно найти в специальной, справочной, учебной и даже научно-популярной литературе [11, 46, 59, 69, 85, 90].

§ 3.5. ПРОСТАЯ ПРОЦЕДУРА ПОСТРОЕНИЯ СИСТЕМЫ ОРТОГОНАЛЬНЫХ МНОГОЧЛЕНОВ

В целях полиномиальной аппроксимации функций может оказаться более удобным использование не конкретных классических ортогональных многочленов, введенных в предыдущем параграфе и жестко привязанных к конкретному скалярному произведению (они еще потребуются в дальнейшем, см. гл. 5), а последовательное построение системы многочленов, взаимно ортогональных в смысле того или иного скалярного произведения $(\cdot, \cdot)$, фиксируемого в конкретных ситуациях.

Итак, будем строить такую систему многочленов $\{q_k\}_{k=0}^m$, что:

$$1) \quad q_0 := 1;$$

- 2) $q_k = q_k(x)$ — многочлен степени k с коэффициентом 1 при старшей степени x ;
- 3) $(q_k, q_j) = 0$ при любых $j \neq k$ (при этом заведомо $(q_k, q_k) = \|q_k\|^2 \neq 0$).

Эта система будет получаться последовательным подсоединением к заданному ее элементу $q_0 := 1$ многочленов повышающихся на единицу степеней, ортогональных всем предыдущим.

Полагая $q_1 := x - \alpha_1$, из условия $q_1 \perp q_0$, т.е. из $(q_1, q_0) = 0$, по свойствам скалярного произведения имеем

$$(x, q_0) - \alpha_1(q_0, q_0) = 0,$$

откуда получаем выражение коэффициента α_1 :

$$\alpha_1 = \frac{(x, q_0)}{(q_0, q_0)}. \quad (3.30)$$

Далее воспользуемся тем фактом, что рассмотренные выше классические ортогональные многочлены удовлетворяют трехчленному рекуррентному соотношению (3.29) с некоторыми наборами параметров $\alpha_n, \beta_n, \gamma_n$. В силу договоренности о том, что здесь будут строиться ортогональные многочлены $q_k(x)$ со старшим членом x^k , можно ограничиться меньшим числом параметров и попытаться находить $(k+1)$ -й многочлен $q_{k+1}(x)$ по уже известным многочленам $q_{k-1}(x)$ и $q_k(x)$ по формуле

$$q_{k+1} = xq_k - \alpha_{k+1}q_k - \beta_kq_{k-1}, \quad (3.31)$$

которая будет полностью определять дальнейший процесс построения $q_2, q_3, q_4, \dots$, если будут известны законы вычисления коэффициентов α_{k+1} и β_k при $k = 1, 2, 3, \dots$.

Сначала убедимся, что формула (3.31) «работает» при $k = 1$. Для этого нужно показать, что в равенстве $q_2 = xq_1 - \alpha_2q_1 - \beta_1q_0$ можно однозначно найти коэффициенты α_2 и β_1 так, чтобы многочлен q_2 был ортогонален одновременно многочленам q_0 и q_1 . Имеем:

$$\begin{cases} (q_2, q_0) = 0, \\ (q_2, q_1) = 0 \end{cases} \Leftrightarrow \begin{cases} (xq_1, q_0) - \alpha_2(q_1, q_0) - \beta_1(q_0, q_0) = 0, \\ (xq_1, q_1) - \alpha_2(q_1, q_1) - \beta_1(q_0, q_1) = 0. \end{cases}$$

Но многочлен q_1 , благодаря (3.30), таков, что $(q_1, q_0) = 0$; поэтому из последней системы сразу получаем

$$\beta_1 = \frac{(xq_1, q_0)}{(q_0, q_0)}, \quad \alpha_2 = \frac{(xq_1, q_1)}{(q_1, q_1)}.$$

Предположим, что уже построены многочлены $q_0, q_1, \dots, q_k$ такие, что имеет место их попарная ортогональность:

$$q_i \perp q_j \quad \forall i, j \in \{0, 1, \dots, k\} : i \neq j. \quad (3.32)$$

Беря многочлен q_{k+1} в форме (3.31), покажем, что он заведомо ортогонален каждому из многочленов от q_0 до q_{k-2} независимо от значений коэффициентов α_{k+1} и β_k , и найдем эти значения из требования ортогональности многочлена q_{k+1} многочленам q_{k-1} и q_k .

Для этого сначала рассматриваем скалярные произведения (q_{k+1}, q_j) при $j = 0, 1, \dots, k-2$:

$$(q_{k+1}, q_j) = (xq_k, q_j) - \alpha_{k+1}(q_k, q_j) - \beta_k(q_{k-1}, q_j).$$

В этом равенстве имеем $(q_k, q_j) = 0$ и $(q_{k-1}, q_j) = 0$, в силу договоренности (3.32), а равенство нулю скалярного произведения (xq_k, q_j) можно установить на основе равенства $(xq_k, q_j) = (q_k, xq_j)$, очевидного, по крайней мере, для используемых здесь скалярных произведений вида (3.7), (3.9) (для которых в выражении скалярного произведения степенных функций (x^k, x^j) применимо их свойство $x^k \cdot x^j = x^{k+j}$). Действительно, так как при $j \leq k-2$ степень многочлена xq_j не превосходит $k-1$ и так как ортогональные многочлены q_j линейно независимы, то многочлен xq_j может быть представлен в виде линейной комбинации многочленов $q_0, q_1, \dots, q_{j+1}$ с некоторыми коэффициентами $\gamma_0, \gamma_1, \dots, \gamma_{j+1}$, и значит,

$$(q_k, xq_j) = \gamma_0(q_k, q_0) + \gamma_1(q_k, q_1) + \dots + \gamma_{j+1}(q_k, q_{j+1}) = 0,$$

поскольку $j+1 \leq k-1$ и выполняется условие (3.32).

Далее, из требования $(q_{k+1}, q_{k-1}) = 0$, т.е. из

$$(xq_k, q_{k-1}) - \alpha_{k+1}(q_k, q_{k-1}) - \beta_k(q_{k-1}, q_{k-1}) = 0,$$

учитывая, что $(q_k, q_{k-1}) = 0$, находим

$$\beta_k = \frac{(xq_k, q_{k-1})}{(q_{k-1}, q_{k-1})}, \quad (3.33)$$

и из требования $(q_{k+1}, q_k) = 0$, т.е. из

$$(xq_k, q_k) - \alpha_{k+1}(q_k, q_k) - \beta_k(q_{k-1}, q_k) = 0,$$

аналогично получаем

$$\alpha_{k+1} = \frac{(xq_k, q_k)}{(q_k, q_k)}. \quad (3.34)$$

Таким образом, завязывая воедино выведенные для выражения (3.31) формулы коэффициентов (3.30), (3.33) и (3.34), приходим к следующей *итерационной процедуре получения ортогональных многочленов все повышающихся степеней*:

$$q_{k+1}(x) = xq_k(x) - \frac{(xq_k(x), q_k(x))}{(q_k(x), q_k(x))} q_k(x) - \frac{(xq_k(x), q_{k-1}(x))}{(q_{k-1}(x), q_{k-1}(x))} q_{k-1}(x), \quad (3.35)$$

$$\text{где } k = 1, 2, \dots; \quad q_0(x) \equiv 1, \quad q_1(x) = x - \frac{(x, q_0(x))}{(q_0(x), q_0(x))}.$$

§ 3.6. АППРОКСИМАЦИЯ ФУНКЦИЙ МНОГОЧЛЕНАМИ ФУРЬЕ

Резюмируя результаты §§ 3.2 и 3.5, приходим к выводу, что построение многочлена m -й степени $Q_m(x)$, осуществляющего *наилучшее среднеквадратическое приближение* для заданной функции $f(x)$, можно выполнить достаточно просто и численно устойчиво, представляя ее отрезком ряда Фурье:

$$f(x) \approx Q_m(x) = \sum_{j=0}^m c_j q_j(x), \quad (3.36)$$

где попарно ортогональные многочлены $q_j(x)$ можно получать итерационным процессом (3.35), а коэффициенты Фурье суть (см.(3.18))

$$c_j = \frac{(q_j(x), f(x))}{(q_j(x), q_j(x))}. \quad (3.37)$$

Те или иные конкретные аппроксимации $f(x)$ типа (3.36) получаются фиксированием тех или иных конкретных скалярных произведений.

Например, аппроксимируя функцию $f(x)$ на отрезке $[a, b]$ многочленом $Q_m(x)$ (3.36) таким образом, чтобы минимизировалось определенное равенством (3.8) среднеквадратическое

отклонение многочлена $Q_m(x)$ от данной функции $f(x)$, в формулах (3.35) и (3.37) нужно использовать интегральное скалярное произведение (3.7) (с множителем $\frac{1}{b-a}$ или без него, что не играет роли). В результате в представлении (3.36) многочлены $q_j(x)$ должны получаться последовательно, начиная с $q_0(x) \equiv 1$,

$$q_1(x) = x - \frac{\int_a^b x dx}{\int_a^b dx} = x - \frac{a+b}{2}, \quad \text{рекуррентным равенством}$$

$$q_{k+1}(x) = xq_k(x) - \frac{\int_a^b xq_k^2(x)dx}{\int_a^b q_k^2(x)dx}q_k(x) - \frac{\int_a^b xq_k(x)q_{k-1}(x)dx}{\int_a^b q_{k-1}^2(x)dx}q_{k-1}(x)$$

при $k = 1, 2, \dots, m-1$, а коэффициенты c_j должны вычисляться по формуле

$$c_j = \frac{\int_a^b q_j(x)f(x)dx}{\int_a^b q_j^2(x)dx} \quad \text{при } j = 0, 1, \dots, m.$$

Рассмотрим теперь случай, когда функция $f(x)$ задана таблицей своих значений в точках $x_0 < x_1 < \dots < x_n$ отрезка $[a, b]$. Для аппроксимации этой сеточной функции многочленом Фурье степени m при $m \leq n$ на основе формул (3.35)–(3.37) используем евклидово скалярное произведение (3.9) (опять-таки с множителем $\frac{1}{n+1}$ или без него), согласно которому, например,

$$(q_0, q_0) = \sum_{i=0}^n 1 = n+1, \quad (x, q_0) = \sum_{i=0}^n x_i, \quad \text{т.е. } \alpha_1 = \frac{1}{n+1} \sum_{i=0}^n x_i, \quad \text{и т.д.}$$

Если при этом степень m аппроксимирующего многочлена $Q_m(x)$ жестко зафиксирована, то этот случай не имеет принципиального отличия от предыдущего. А именно, начиная с

$$q_0(x) \equiv 1, \quad q_1(x) = x - \frac{1}{n+1} \sum_{i=0}^n x_i, \quad c_0 = \frac{1}{n+1} \sum_{i=0}^n f(x_i), \quad (3.38)$$

при $k = 1, 2, \dots, m-1$ вычисляем коэффициенты

$$\alpha_{k+1} = \frac{\sum_{i=0}^n x_i q_k^2(x_i)}{\sum_{i=0}^n q_k^2(x_i)}, \quad \beta_k = \frac{\sum_{i=0}^n x_i q_k(x_i) q_{k-1}(x_i)}{\sum_{i=0}^n q_{k-1}^2(x_i)} \quad (3.39)$$

и с их помощью конструируем многочлены $q_2(x)$, $q_3(x)$, ..., $q_m(x)$ по формуле

$$q_{k+1}(x) = xq_k(x) - \alpha_{k+1}q_k(x) - \beta_k q_{k-1}(x); \quad (3.40)$$

затем при $j = 1, 2, \dots, m$ вычисляем коэффициенты Фурье

$$c_j = \frac{\sum_{i=0}^n q_j(x_i) f(x_i)}{\sum_{i=0}^n q_j^2(x_i)} \quad (3.41)$$

для их подстановки в выражение (3.36).

Заметим, что при $m = n$ получаемый таким образом многочлен Фурье $Q_n(x)$ наилучшего среднеквадратического приближения сеточной функции $f(x)$ на $[a, b]$ должен совпасть с соответствующим интерполяционным многочленом Лагранжа $L_n(x)$ (см (1.6)). Следовательно, *многочлен*

$$Q_n(x) = \sum_{j=0}^n c_j q_j(x)$$

с определенными с помощью формул (3.38)–(3.41) коэффициентами c_j и ортогональными многочленами $q_j(x)$ дает еще одну форму представления интерполяционного многочлена наряду с несколькими другими, изученными в гл. 1.

Пусть теперь речь идет не о представлении функции $f(x)$, заданной таблицей своих приближенных значений $f(x_i)$ ($i = 0, 1, \dots, n$), многочленом $Q_m(x)$ определенной степени m , а о вычислении ее приближенного значения $f(\tilde{x})$ в заданной точке $\tilde{x} \in [a, b]$, причем считается, что $a \approx x_0 < x_1 < \dots < x_n \approx b$ и $\tilde{x} \neq x_i$.

Учитывая, что повышение степени m многочлена Фурье $Q_m(x)$, аппроксимирующего $f(x)$ на $[a, b]$, улучшает качество аппроксимации, т. е. уменьшает величину среднеквадратической ошибки (3.16) до тех пор, пока на погрешность метода не станет существенно влиять погрешность округлений (вычислений), *есть смысл совместить процесс вычисления коэффициентов Фурье с процедурой ортогонализации* в одном цикле и производить последовательное вычисление значений $Q_0(\tilde{x})$, $Q_1(\tilde{x})$, ..., все более точно приближающих значение $f(\tilde{x})$. Алгоритм, реализующий такой подход, должен предусматривать вычисление

на каждом k -м шаге среднеквадратических отклонений

$$\rho_k = \sqrt{\frac{1}{n+1} \sum_{i=0}^n (f(x_i) - Q_k(x_i))^2}$$

и проверку их на убывание при переходе от k к $k+1$, проверку величин $Q_k(\tilde{x})$ на сближение при увеличении k (т.е. того, что $|Q_{k+1}(\tilde{x}) - Q_k(\tilde{x})| < |Q_k(\tilde{x}) - Q_{k-1}(\tilde{x})|$), а также того, что $k < n$.

Кстати, при небольшом числе узлов, т.е. при небольших значениях n , случай $k = n$ можно использовать для проверки правильности работы алгоритма. Действительно, в этом случае многочлен $Q_k(x) \equiv Q_n(x)$ должен стать интерполяционным, но поскольку он строился из других соображений, то сравнение величины $f(x_i) - Q_n(x_i)$ или ρ_n с нулем не тривиально (не то, что при классической интерполяции).

УПРАЖНЕНИЯ

3.1. Найдите нормальное псевдорешение системы

$$\begin{cases} 2x_1 + 3x_2 - 4x_3 = 1.2, \\ 3x_1 + x_2 + 7x_3 = 4.7, \\ x_1 - 4x_2 - 3x_3 = 6.1, \\ 4x_1 + 3x_2 - 5x_3 = 5.2, \\ 2x_1 - 4x_2 + 3x_3 = 7.9. \end{cases}$$

3.2. Для функции $y = f(x)$, заданной таблицей

x	0.5	1	1.5	2	2.5
y	10.5	1.6	0.55	0.26	0.15

подберите подходящий вид аппроксимирующей ее нелинейной зависимости из следующих:

а) $y = a \cdot x^b$; б) $y = a \cdot e^{bx}$; в) $y = a + \frac{b}{x}$; г) $y = \frac{1}{a + bx}$,

находя методом наименьших квадратов их параметры и сравнивая между собой среднеквадратические погрешности.

3.3. Постройте наилучшие среднеквадратические линейные аппроксимации для функций

а) $f(x) = \sqrt{x}$ при $x \in [0, 1]$,

$$\text{б) } f(x) = \frac{1}{x} \quad \text{при } x \in [1, 2],$$

$$\text{в) } f(x) = \ln(1+x) \quad \text{при } x \in [0, 1]$$

непосредственным применением метода наименьших квадратов. Сравните полученные результаты с результатами чебышевской аппроксимации (см. упр.2.4):

А) по величине среднеквадратической погрешности;

Б) по величине максимальной погрешности на заданном отрезке.

3.4. Функция $y = f(x)$ задана таблицей

x	0	0.1	0.2	0.3	0.4	0.5	0.6	0.8	0.9	1.0
y	0	0.095	0.182	0.262	0.337	0.406	0.470	0.588	0.642	0.693

Методом наименьших квадратов аппроксимируйте данную функцию $f(x)$ функцией $\varphi(x) = a_0 + a_1x$ по трем точкам ($x = 0$, $x = 0.5$ и $x = 1.0$), по шести ($x = 0, 0.2, 0.4, 0.6, 0.8, 1.0$) и по всем одиннадцати точкам. Сравните результаты (в разных метриках) с результатом применения интегрального МНК, зная, что $f(x) = \ln(1+x)$ (см. упр. 3.3в)).

3.5. Покажите, что матрица системы (3.22) при $x_i = \frac{i}{n} \in [0, 1]$, $m = 1$ и $n \rightarrow \infty$ совпадает с матрицей Гильберта H_2 .

3.6. Для функций а)–в) упр.3.3 на заданных отрезках постройте многочлены Фурье первой и второй степени. Найдите среднеквадратические ошибки результатов. Совпадают ли построенные многочлены первой степени с линейными функциями, полученными в упр. 3.3?

ГЛАВА 4 || ИНТЕРПОЛЯЦИОННЫЕ СПЛАЙНЫ^{*)}

Рассматривается высокотехнологичный способ приближения таблично заданных функций с помощью составных функций, звеньями которых служат многочлены небольших степеней, допускающие гладкую стыковку. Наряду со ставшими уже классическими естественными кубическими сплайнами, здесь изучаются интерполяционные параболические (квадратичные) сплайны, а также базисные сплайны (иначе, *B-сплайны*) и эрмитовы (локальные) сплайны нечетных и четных степеней. Дается представление о сглаживании эмпирических данных посредством простейших линейных фильтров и указывается на возможность построения сглаживающих сплайнов.

§ 4.1. КУСОЧНО-ПОЛИНОМИАЛЬНАЯ АППРОКСИМАЦИЯ. ЛИНЕЙНЫЕ ФИЛЬТРЫ

В тех случаях, когда промежуток $[a, b]$, на котором нужно подменить функцию $f(x)$ функцией $\varphi(x)$, велик, и отсутствуют основания считать данную функцию $f(x)$ достаточно гладкой при $x \in [a, b]$, нет смысла пытаться повышать качество ее полиномиальной аппроксимации за счет использования в роли $\varphi(x)$ многочленов высоких степеней. Более перспективным в этих условиях является применение *кусочно-полиномиальной аппроксимации* $f(x)$, предполагающей, что аппроксимирующая функция $\varphi(x)$ составляется из отдельных многочленов, как правило, одинаковой небольшой степени, определенных каждый на своей части отрезка $[a, b]$. При этом, если функция $f(x)$ непрерывна и имеется достаточное количество точечной информации о ней, то можно рассчитывать приблизить ее на $[a, b]$ сколь угодно хорошо кусочно-полиномиальной функцией $\varphi(x)$ только за счет увеличения числа частичных промежутков, составляющих $[a, b]$,

^{*)} Прочитируем абзац, с которого начинается посвященный сплайнам параграф в книге [9]:

«Способ приближения сплайнами интересен, кроме всего прочего, отношением к нему специалистов. Одни считают его универсальным методом решения проблем, стоящих перед численным анализом, и ищут применения ему в самых различных направлениях. Другие рассматривают его как очередную дань переменчивой моде. По-видимому, истина находится где-то посередине».

при любых фиксированных степенях составных многочленов и любых способах согласования $f(x)$ и $\varphi(x)$.

Использование низких степеней многочленов, составляющих $\varphi(x)$, позволяет легко находить их коэффициенты как из интерполяционных, так и из иных условий.

Так, если заданы значения y_i функции $y = f(x)$ на системе узлов x_i таких, что

$$a \leq x_0 < x_1 < \dots < x_n \leq b, \quad (4.1)$$

и требуется аппроксимировать $f(x)$ *кусочно-линейной функцией* $\varphi(x)$, исходя из условий интерполяции

$$\varphi(x_i) = y_i, \quad (i = 0, 1, \dots, n),$$

то, беря функцию $\varphi(x)$ в виде

$$\varphi(x) = \begin{cases} a_1x + b_1 & \text{при } x \in [x_0, x_1], \\ a_2x + b_2 & \text{при } x \in [x_1, x_2], \\ \dots & \dots \\ a_nx + b_n & \text{при } x \in [x_{n-1}, x_n], \end{cases} \quad (4.2)$$

для нахождения n пар ее коэффициентов a_k, b_k ($k = 1, 2, \dots, n$) имеем систему из $2n$ линейных уравнений

$$\begin{cases} \begin{cases} a_1x_0 + b_1 & = y_0, \\ a_1x_1 + b_1 & = y_1; \end{cases} \\ \begin{cases} a_2x_1 + b_2 & = y_1, \\ a_2x_2 + b_2 & = y_2; \end{cases} \\ \dots \\ \begin{cases} a_nx_{n-1} + b_n & = y_{n-1}, \\ a_nx_n + b_n & = y_n, \end{cases} \end{cases} \quad (4.3)$$

причем, как видим, каждая пара соседних уравнений системы (4.3), имеющих коэффициенты с одинаковыми индексами, не связана с остальными и может решаться отдельно.

Аналогично, каждое звено *кусочно-квадратичной функции* (при $n = 2m$ в (4.1))

$$\varphi(x) = \begin{cases} a_1x^2 + b_1x + c_1 & \text{при } x \in [x_0, x_2], \\ a_2x^2 + b_2x + c_2 & \text{при } x \in [x_2, x_4], \\ \dots & \dots \\ a_mx^2 + b_mx + c_m & \text{при } x \in [x_{2m-2}, x_{2m}] \end{cases} \quad (4.4)$$

определяется тройкой коэффициентов a_k, b_k, c_k ($k = 1, 2, \dots, m$), которые могут быть найдены последовательным решением (при $k = 1, 2, \dots, m$) трехмерных линейных систем

$$\begin{cases} a_k x_{2k-2}^2 + b_k x_{2k-2} + c_k = y_{2k-2}, \\ a_k x_{2k-1}^2 + b_k x_{2k-1} + c_k = y_{2k-1}, \\ a_k x_{2k}^2 + b_k x_{2k} + c_k = y_{2k}, \end{cases} \quad (4.5)$$

соответствующим выставленным интерполяционным условиям.

Фактически, в рассмотренных случаях речь идет о последовательной линейной интерполяции (1.7) по перемещаемым вдоль отрезка $[a, b]$ парам соседних точек разбиения (4.1) и о последовательной квадратичной интерполяции (1.8) по тройкам таких точек.

Пример 4.1. Для функции $y = f(x)$, заданной таблицей

x	0	0.5	1	2	3	4	5
$f(x)$	1.5	0	0	2	2	1	2

выполним простейшие кусочно-линейное и кусочно-квадратичное интерполирования.

Осуществляя линейное интерполирование данной функции на каждом из элементарных промежутков, определяемых соседними числами верхней строки таблицы, получаем, что можно считать $f(x) \approx \varphi_1(x)$, где

$$\varphi_1(x) = \begin{cases} -3x + 1.5, & x \in [0, 0.5], \\ 0, & x \in [0.5, 1], \\ 2x - 2, & x \in [1, 2], \\ 2, & x \in [2, 3], \\ -x + 5, & x \in [3, 4], \\ x - 3, & x \in [4, 5]. \end{cases}$$

Квадратичное интерполирование по тройкам известных точек отрезков $[0, 1]$, $[1, 3]$ и $[3, 5]$ приводит к приближенному равенству $f(x) \approx \varphi_2(x)$, где

$$\varphi_2(x) = \begin{cases} 3x^2 - 4.5x + 1.5, & x \in [0, 1], \\ -x^2 + 5x - 4, & x \in [1, 3], \\ x^2 - 8x + 17, & x \in [3, 5]. \end{cases}$$

Графики функций $\varphi_1(x)$ и $\varphi_2(x)$ показаны на рис. 4.1 и 4.2.

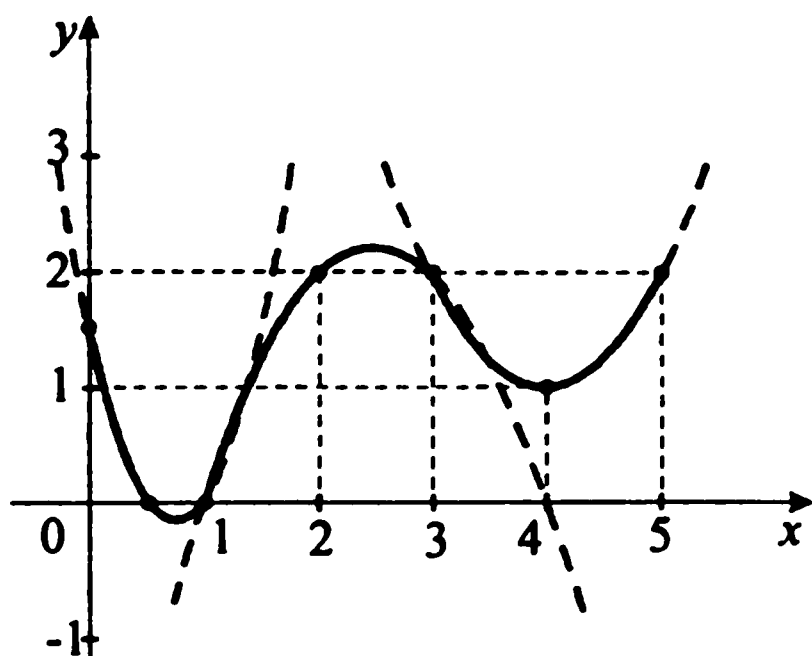
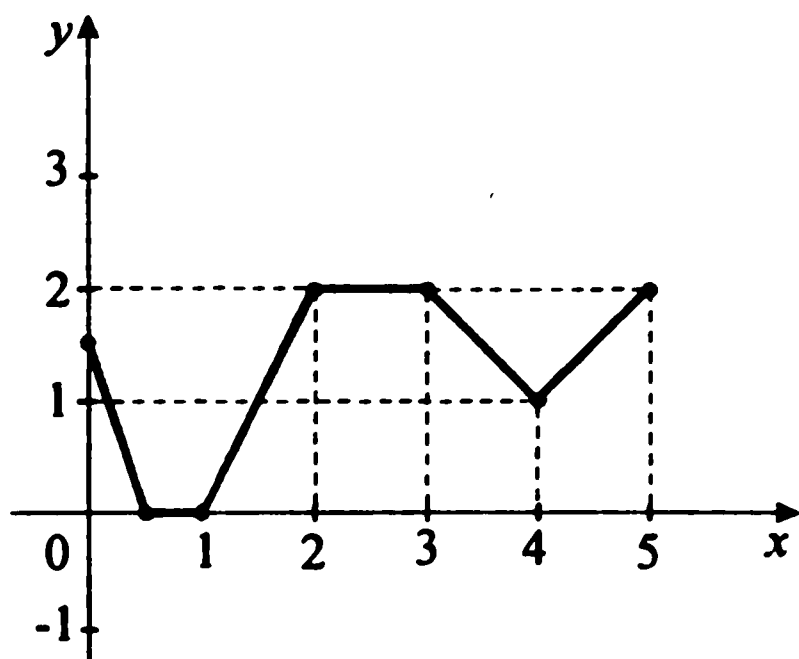


Рис. 4.1. График функции $y = \varphi_1(x)$

Рис. 4.2. График функции $y = \varphi_2(x)$

Кусочно-линейная аппроксимация таблично заданных функций, проводимая на основе метода наименьших квадратов, приводит к понятию линейного фильтра.

Предположим, что функция $f(x)$ задается «длинной» таблицей своих значений $f_i \approx f(x_i)$ на системе равноотстоящих точек $x_i = x_0 + ih$, и пусть известно, что некоторые из значений f_i могут содержать достаточно большие случайные ошибки («выбросы»). Тогда, прежде чем производить какую-либо содержательную математическую обработку таких данных (например, полученных из эксперимента), целесообразно сначала произвести их *сглаживание*, чтобы уменьшить роль выбросов. Рассмотрим вкратце один из простейших подходов к такому сглаживанию.

Если при кусочно-линейной интерполяции линейная функция на элементарном отрезке $[x_{i-1}, x_i]$ однозначно определялась по двум точкам из условия совпадения ее значений в них со значениями f_{i-1} и f_i данной функции $f(x)$, то здесь на каждом из отрезков $[x_{i-1}, x_{i+1}]$ длиной $2h$ функция $f(x)$ будет подменяться наилучшим линейным среднеквадратическим приближением $\varphi_i(x)$, построенным по трем значениям f_{i-1} , f_i и f_{i+1} , и отфильтрованным значением в точке x_i будет считаться значение $\varphi_i(x_i)$.

Итак, на первом этапе ставим задачу: для функции $f(x)$, представленной таблицей

$$\begin{array}{c|c|c|c|c} x & x_0 & x_1 & \dots & x_n \\ \hline f(x) & f_0 & f_1 & \dots & f_n \end{array}, \quad (4.6)$$

где $x_i = x_0 + ih$, $i = 0, 1, \dots, n$, найти такую функцию $\varphi(x)$, составленную из линейных функций

$$\varphi_i(x) := a_i + b_i(x - x_i)$$

(с перекрытием), чтобы

$$f(x) \approx \varphi_i(x) \quad \forall x \in [x_{i-1}, x_{i+1}]$$

в смысле

$$\sum_{k=i-1}^{i+1} (f_k - \varphi_i(x_k))^2 = \min.$$

Эта знакомая по § 3.1 задача нахождения коэффициентов a_i , b_i методом наименьших квадратов решается непосредственно. Имеем:

$$(f_{i-1} - a_i + b_i h)^2 + (f_i - a_i)^2 + (f_{i+1} - a_i - b_i h)^2 = \min$$

$$\Leftrightarrow \begin{cases} (f_{i-1} - a_i + b_i h) + (f_i - a_i) + (f_{i+1} - a_i - b_i h) = 0, \\ (f_{i-1} - a_i + b_i h)h - (f_{i+1} - a_i - b_i h)h = 0. \end{cases}$$

Отсюда получаем $a_i = \frac{1}{3}(f_{i-1} + f_i + f_{i+1})$, $b_i = \frac{1}{2h}(f_{i+1} - f_{i-1})$.

Следовательно, на каждом из промежутков $[x_{i-1}, x_{i+1}]$ функция $f(x)$ может быть подменена линейной функцией

$$\varphi_i(x) = \frac{1}{3}(f_{i-1} + f_i + f_{i+1}) + \frac{f_{i+1} - f_{i-1}}{2h}(x - x_i).$$

Варьированием i от 1 до $n-1$ для функции (4.6) таких линейных функций $\varphi_i(x)$ будет построено $n-1$; на каждый элементарный промежуток $[x_i, x_{i+1}]$ (кроме первого и последнего) их придется по две.

Второй этап состоит в пересчете данной таблицы (4.6) заменой значения f_i на значение

$$\varphi_i(x_i) = \frac{1}{3}(f_{i-1} + f_i + f_{i+1})$$

при каждом $i = 1, 2, \dots, n-1$. Доопределив новую табличную функцию, например, значениями $\varphi_0(x_0) := f_0$ и $\varphi_n(x_n) := f_n$, получаем вместо (4.6) табличную зависимость

x	x_0	x_1	$\dots$	x_{n-1}	x_n
$\varphi(x)$	$\varphi_0(x_0)$	$\varphi_1(x_1)$	$\dots$	$\varphi_{n-1}(x_{n-1})$	$\varphi_n(x_n)$

в которой «в целом» сохраняется характер поведения исходной функции $f(x)$ и уменьшена роль ее отдельных значений.

Описанная процедура называется *осреднением по трем точкам* и является простым частным случаем *линейного фильтра*. Более глубокое осмысление роли линейных фильтров с помощью гармонического анализа можно найти в теории цифровой обработки сигналов; некоторые элементы такого анализа роли фильтров см., например, в [81, 90].

Отметим, что осреднение проводят и по большему числу точек, а иногда прибегают к двойной фильтрации табличных данных; правда, при глубокой фильтрации есть риск потерять полезную информацию о таблично заданной функции.

Пример 4.2. Функция $y = f(x)$, заданная таблицей

x	1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2
$f(x)$	0.0	0.1	0.2	0.3	1.0	0.4	0.5	0.5	0.1	0.6	0.7

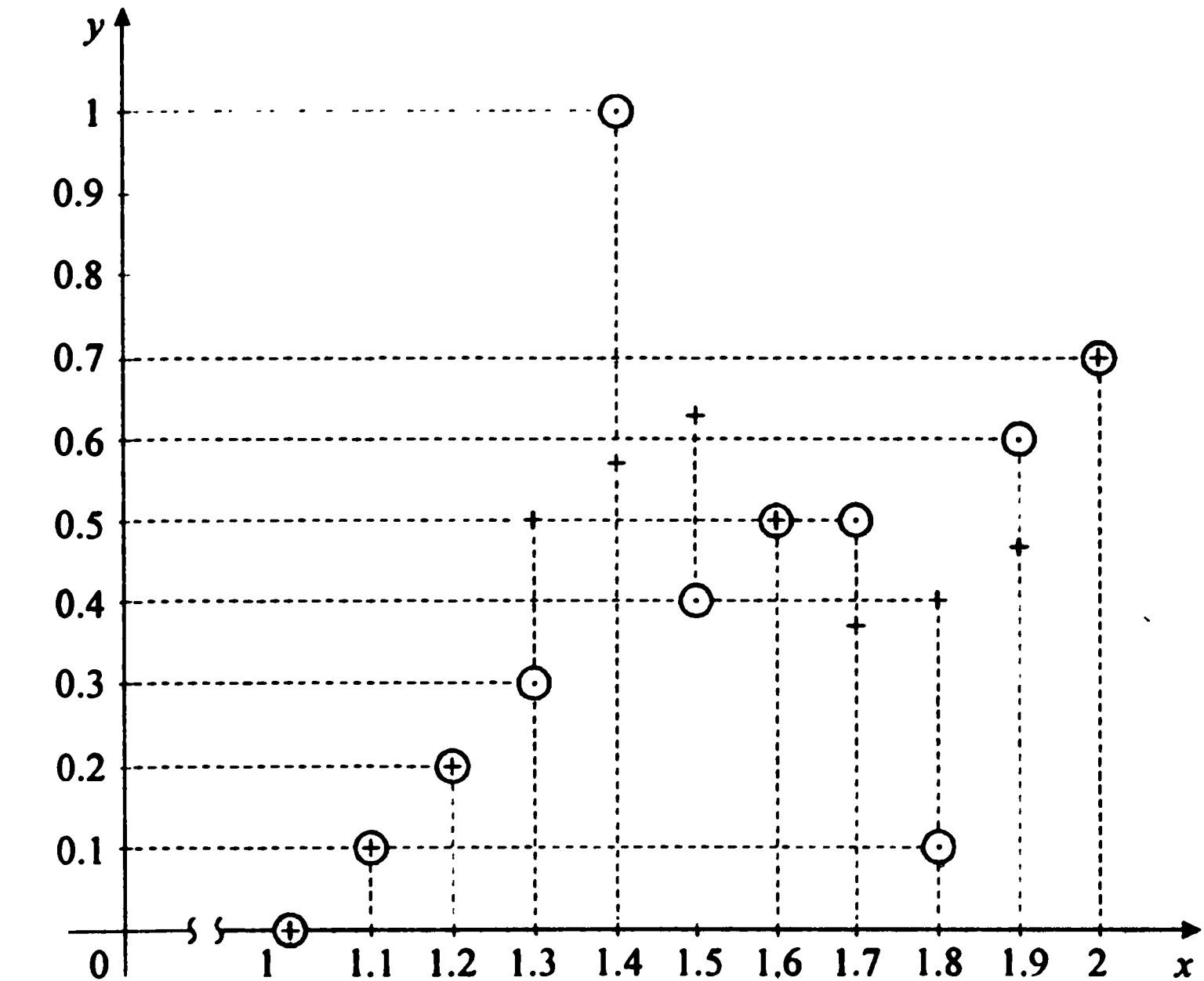


Рис. 4.3. Результат осреднения функции по трем точкам

представляет собой сеточную функцию $y = \ln x$ с систематическими ошибками округления в каждой точке сетки, не превосходящими 0.05, и двумя выбросами в точках $x = 1.4$ и $x = 1.8$. Осреднение по трем соседним

точкам, т.е. замена каждого значения данной табличной функции средним арифметическим трех ближайших значений, приводит к таблице

x	1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2
$\varphi(x)$	0.0	0.1	0.2	0.50	0.57	0.63	0.5	0.37	0.40	0.47	0.7

(где значения, подвергшиеся коррекции, записаны с двумя знаками после запятой). Эффект сглаживания, полученный таким простым фильтром, можно увидеть на рис. 4.3, где данные приближенные значения $f(x_i)$ помечены кружочками, а осредненные значения $\varphi(x_i)$ — крестиками.

§ 4.2. ОПРЕДЕЛЕНИЕ СПЛАЙНА. ИНТЕРПОЛЯЦИОННЫЙ КУБИЧЕСКИЙ СПЛАЙН ДЕФЕКТА 1

Пусть на отрезке $[a, b]$ задана упорядоченная система несовпадающих точек x_k ($k = 0, 1, \dots, n$).

Определение 4.1. *Сплайном $S_m(x)$ называется определенная на $[a, b]$ функция, принадлежащая классу $C^l[a, b]$ l раз непрерывно дифференцируемых функций, такая, что на каждом промежутке $[x_{k-1}, x_k]$ ($k = 1, 2, \dots, n$) — это многочлен m -й степени. Разность $d := m - l$ между степенью сплайна m и показателем его гладкости l называется **дефектом** сплайна.*

Если сплайн $S_m(x)$ строится по некоторой функции $f(x)$ так, чтобы выполнялись условия $S_m(x_i) = f(x_i)$, то такой сплайн называется **интерполяционным сплайном** для функции $f(x)$; при этом **узлы сплайна** x_k , вообще говоря, могут не совпадать с узлами интерполяции x_i .

Тривиальные примеры интерполяционных сплайнов можно найти в предыдущем параграфе: кусочно-линейная функция $\varphi(x)$, определенная в (4.2) с параметрами a_k, b_k , удовлетворяющими условиям (4.3), очевидно, является интерполяционным сплайном степени 1 дефекта 1, а кусочно-квадратичная функция (4.4) при условиях (4.5) есть интерполяционный сплайн степени 2 дефекта 2.

Совпадение дефекта сплайна с его степенью обеспечивает просто непрерывность сплайна. Интерес представляет построение сплайнов с большей гладкостью, т. е. с малым дефектом. Такие сплайны являют собой дальнейшее совершенствование идеи кусочно-полиномиальной аппроксимации.

Рассмотрим наиболее известный и широко применяемый интерполяционный сплайн степени 3 дефекта 1. При этом будем исходить из предположения, что узлы сплайна

$$a = x_0, x_1, x_2, \dots, x_{n-1}, x_n = b \quad (4.7)$$

одновременно служат узлами интерполяции, т.е. в них известны значения функции $f_k := f(x_k)$, $k = 0, 1, \dots, n$.

Определение 4.2. Кубическим сплайном дефекта 1, интерполирующим на отрезке $[a, b]$ данную функцию $f(x)$, называется функция

$$g(x) := \left\{ g_k(x) := a_k + b_k(x - x_k) + c_k(x - x_k)^2 + d_k(x - x_k)^3 \right. \\ \left. \text{при } x \in [x_{k-1}, x_k] \right\}_{k=1}^n, \quad (4.8)$$

удовлетворяющая совокупности условий:

- а) $g(x_k) = f_k$ (условие интерполяции в узлах сплайна);
- б) $g(x) \in C_{[a, b]}^2$ (двойная непрерывная дифференцируемость);
- в) $g''(a) = g''(b) = 0$ (краевые условия).

Определенный таким образом сплайн называют еще *естественным* или *чертежным сплайном*^{*} и связано это со следующим обстоятельством. Желая провести плавную линию через заданные точки плоскости, чертежники фиксировали в этих точках гибкую упругую рейку, тогда под влиянием упругих сил она принимала нужную форму, обеспечивающую минимум потенциальной энергии. Несложно убедиться, что определяемая условиями а)–в) функция (4.8), представляющая собой кубический n -звенник с гладким сопряжением звеньев, служит математическим описанием такого чертежного приема.

С этой целью достаточно показать, что значение функционала $\Phi(g) := \int_a^b (g''(x))^2 dx$, характеризующее указанную величину потенциальной энергии закрепленной в $(n+1)$ -й точке (4.7) упругой рейки, не превосходит величины $\Phi(f) := \int_a^b (f''(x))^2 dx$,

^{*} Английское слово *spline* переводится как (гибкая) планка, рейка, брусок.

соответствующей потенциальной энергии закрепленной в тех же точках рейки, но принимающей любую другую, отличную от $y = g(x)$ форму $y = f(x)$.

Более удобно это сделать, рассматривая $\Phi(f - g)$. Имеем:

$$\Phi(f - g) = \int_a^b (f''(x) - g''(x))^2 dx = \Phi(f) + \Phi(g) - 2 \int_a^b f''(x)g''(x)dx =$$

(прибавим и вычтем величину $2\Phi(g)$)

$$= \Phi(f) - \Phi(g) - 2 \int_a^b (f''(x) - g''(x))g''(x)dx =$$

(применим правило интегрирования «по частям»)

$$= \Phi(f) - \Phi(g) - 2(f'(x) - g'(x))g''(x) \Big|_a^b + 2 \int_a^b (f'(x) - g'(x))g'''(x)dx =$$

(при подстановке границ учтем краевые условия в), а к последнему интегралу применим свойство аддитивности по промежутку интегрирования)

$$= \Phi(f) - \Phi(g) + 2 \sum_{k=1}^n \int_{x_{k-1}}^{x_k} (f'(x) - g'_k(x))g'''_k(x)dx =$$

$$= \Phi(f) - \Phi(g) + 12 \sum_{k=1}^n d_k [f(x) - g_k(x)]_{x_{k-1}}^{x_k} =$$

(используем условия интерполяции, согласно которым должно быть $g(x_k) = f(x_k)$ и $g(x_{k-1}) = f(x_{k-1})$)

$$= \Phi(f) - \Phi(g).$$

Таким образом, $\Phi(g) = \Phi(f) - \Phi(f - g)$, откуда, в силу неотрицательности функционала Φ , следует $\Phi(g) \leq \Phi(f)$.

Замечание 4.1. Краевые условия в определении 4.2 могут быть заменены на другие. Например, можно наложить дополнительные условия на первую производную функции $g(x)$ в точках a и b . В таком случае кубический сплайн (4.8), оставаясь интерполяционным дефекта 1, утрачивает свойство быть естественным.

Для построения по данной функции $f(x)$ интерполирующего ее сплайна (4.8) нужно найти $4n$ его коэффициентов a_k , b_k , c_k , d_k ($k = 1, 2, \dots, n$). Чтобы понять, имеется ли для этого достаточное количество связей, расшифруем фигурирующие в определении 4.2 условия а)–в) через функции $g_k(x)$, составляющие $g(x)$, имея в виду, что в любом внутреннем узле должны

совпадать значения двух соседних звеньев сплайна и двух их первых производных (см. наглядную структуру сплайна на рис. 4.4).

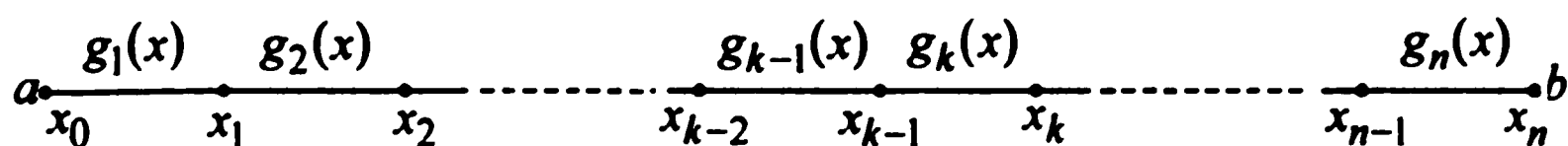


Рис. 4.4. Расположение узлов и звеньев кубического сплайна (4.8)

Имеем:

из условий интерполяции а):

$$g_1(x_0) = f_0, \quad g_k(x_k) = f_k \text{ при } k = 1, 2, \dots, n;$$

из условий гладкой стыковки звеньев сплайна б):

$$\left. \begin{aligned} g_{k-1}(x_{k-1}) &= g_k(x_{k-1}), \\ g'_{k-1}(x_{k-1}) &= g'_k(x_{k-1}), \\ g''_{k-1}(x_{k-1}) &= g''_k(x_{k-1}) \end{aligned} \right\} \text{ при } k = 2, 3, \dots, n;$$

из краевых условий в):

$$g''_1(x_0) = 0, \quad g''_n(x_n) = 0.$$

Как видим, условий оказалось $4n$ — ровно столько, сколько в записи сплайна (4.8) неизвестных коэффициентов. Подставляя сюда выражения функций

$$g_k(x) := a_k + b_k(x - x_k) + c_k(x - x_k)^2 + d_k(x - x_k)^3$$

и их производных

$$g'_k(x) = b_k + 2c_k(x - x_k) + 3d_k(x - x_k)^2 \quad (4.9)$$

и

$$g''_k(x) = 2c_k + 6d_k(x - x_k)$$

через коэффициенты a_k, b_k, c_k, d_k при указанных значениях k и полагая для краткости

$$h_k := x_k - x_{k-1}, \quad (4.10)$$

получаем детализированную систему связей

$$\begin{cases} a_1 - b_1 h_1 + c_1 h_1^2 - d_1 h_1^3 = f_0, & (4.11) \end{cases}$$

$$a_k = f_k \quad \text{при } k = 1, 2, \dots, n, \quad (4.12)$$

$$\begin{cases} a_{k-1} = a_k - b_k h_k + c_k h_k^2 - d_k h_k^3, & (4.13) \end{cases}$$

$$\begin{cases} b_{k-1} = b_k - 2c_k h_k + 3d_k h_k^2, & \text{при } k = 2, 3, \dots, n \quad (4.14) \end{cases}$$

$$\begin{cases} c_{k-1} = c_k - 3d_k h_k, & (4.15) \end{cases}$$

$$c_1 - 3d_1 h_1 = 0, \quad (4.16)$$

$$c_n = 0. \quad (4.17)$$

Теперь ставим задачу выявления эффективного способа нахождения коэффициентов сплайна a_k, b_k, c_k, d_k ($k = 1, 2, \dots, n$) из этой линейной относительно них системы (4.11)–(4.17). С этой целью будем исключать из системы неизвестные a_k, d_k, b_k и сводить все к решению системы относительно неизвестных c_k . При этом для упрощения записей используем уже применявшееся ранее (см. (1.39) в § 1.6) обозначение разделенной разности

$$f(x_{k-1}; x_k) := \frac{f_k - f_{k-1}}{h_k} \quad (4.18)$$

и, кроме того, введем фиктивный коэффициент

$$c_0 = 0. \quad (4.19)$$

Итак, согласно (4.12), коэффициенты a_k известны и равны f_k при любом $k \in \{1, 2, \dots, n\}$. Подставляя их значения в равенства (4.11) и (4.13), приходим к равенству

$$b_k h_k - c_k h_k^2 + d_k h_k^3 = f_k - f_{k-1},$$

справедливому при $k = 1, 2, \dots, n$, откуда с учетом обозначения (4.18) получаем выражение

$$b_k = f(x_{k-1}; x_k) + c_k h_k - d_k h_k^2 \quad (k = 1, 2, \dots, n). \quad (4.20)$$

Далее, из (4.15) и (4.16), если учесть (4.19), можно однозначно выразить d_k через c_k :

$$d_k = \frac{c_k - c_{k-1}}{3h_k} \quad (k = 1, 2, \dots, n). \quad (4.21)$$

С помощью (4.21) избавляемся от d_k в (4.20):

$$b_k = f(x_{k-1}; x_k) + \frac{2}{3} h_k c_k + \frac{1}{3} h_k c_{k-1} \quad (k = 1, 2, \dots, n). \quad (4.22)$$

Теперь пользуемся связью (4.14), подставляя туда b_{k-1} , b_k из (4.22) и d_k из (4.21):

$$\begin{aligned} f(x_{k-2}; x_{k-1}) + \frac{2}{3}h_{k-1}c_{k-1} + \frac{1}{3}h_{k-1}c_{k-2} = \\ = f(x_{k-1}; x_k) + \frac{2}{3}h_k c_k + \frac{1}{3}h_k c_{k-1} - 2h_k c_k + 3h_k^2 \cdot \frac{c_k - c_{k-1}}{3h_k}. \end{aligned}$$

После упрощения отсюда получаем относительно неизвестных c_k трехточечное разностное уравнение второго порядка

$$\begin{aligned} h_{k-1}c_{k-2} + 2(h_{k-1} + h_k)c_{k-1} + h_k c_k = \\ = 3f(x_{k-1}; x_k) - 3f(x_{k-2}; x_{k-1}), \quad (4.23) \end{aligned}$$

где $k = 2, 3, \dots, n$, $c_0 = 0$ (согласно (4.19)) и $c_n = 0$ (согласно (4.17)).

К такому же итогу можно прийти и иными путями [71, 73, 81, 84 и др.].

Так как $|2h_{k-1} + 2h_k|$ заведомо больше, чем $|h_{k-1}| + |h_k|$, то для разностного уравнения (4.23) (по другой терминологии, ленточной системы линейных алгебраических уравнений с трехдиагональной матрицей коэффициентов) выполняется достаточное условие однозначной разрешимости, т.е. существует единственный набор коэффициентов $c_1, c_2, \dots, c_n$, удовлетворяющий (4.23). Найдя эти коэффициенты, все остальные коэффициенты сплайна (4.8) так же однозначно получаем по формулам (4.12), (4.22) и (4.21). Таким образом, справедлива следующая теорема.

Теорема 4.1. При заданных в точках a и b краевых условиях ($g''(a) = g''(b) = 0$) по заданным в узлах

$$a = x_0, x_1, x_2, \dots, x_{n-1}, x_n = b$$

значениям f_k ($k = 0, 1, \dots, n-1$) функции $f(x)$ можно построить единственный интерполирующий ее кубический сплайн $g(x)$ дефекта 1.

Следует отметить высокую технологичность процесса сплайн-интерполирования.

Действительно, упомянутое выше диагональное преобладание в трехдиагональной матрице системы (4.23) обеспечивает корректность и устойчивость **метода прогонки** [2, 18, 21, 80]. Применение этого метода для решения системы (4.23) сводится к вычислению прогоночных коэффициентов по формулам

прямой прогонки

$$\delta_1 = -\frac{h_2}{2h_1 + 2h_2}, \quad \lambda_1 = \frac{3f(x_1; x_2) - 3f(x_0; x_1)}{2h_1 + 2h_2}, \quad (4.24)$$

$$\delta_{k-1} = -\frac{h_k}{2h_{k-1} + 2h_k + h_{k-1}\delta_{k-2}}, \quad (4.25)$$

$$\lambda_{k-1} = \frac{3f(x_{k-1}; x_k) - 3f(x_{k-2}; x_{k-1}) - h_{k-1}\lambda_{k-2}}{2h_{k-1} + 2h_k + h_{k-1}\delta_{k-2}} \quad (4.26)$$

при $k = 3, 4, \dots, n$, а затем к получению искоемых значений c_k обратной прогонкой по формуле

$$c_{k-1} = \delta_{k-1}c_k + \lambda_{k-1}, \quad (4.27)$$

полагая в ней $k = n, n-1, \dots, 2$ и учитывая, что $c_n = 0$. После этого, как уже говорилось, остается подставить c_k в выражения d_k (4.21) и b_k (4.22), а в качестве a_k в (4.8) взять значения f_k . Все вычислительные затраты на построение n -звенного естественного сплайна составят, очевидно, $O(n)$ арифметических операций.

Факт и скорость сходимости сплайн-интерполяционного процесса характеризует следующее приводимое здесь без доказательства утверждение [1, 13, 63, 80, 84]

Теорема 4.2. Пусть $g(x)$ — кубический сплайн дефекта 1, интерполирующий на системе узлов (4.7) отрезка $[a, b]$ четырежды непрерывно дифференцируемую на нем функцию $f(x)$. Тогда при любом фиксированном n найдется такая постоянная $C > 0$, что для любого $x \in [a, b]$

$$|f(x) - g(x)| \leq C \cdot \Delta^4, \quad (4.28)$$

где $\Delta := \max_{1 \leq k \leq n} (x_k - x_{k-1})$.

Все расчетные формулы упрощаются в частном случае, когда сплайн $g(x)$ строится по системе равноотстоящих узлов. Выпишем всю совокупность формул в естественном для вычисления коэффициентов сплайна (4.8) порядке, заменяя в соответствующих равенствах (4.24)–(4.26), (4.22), (4.21) переменный шаг h_k (4.10) на постоянный шаг h . При этом вместо разделенных разностей (4.18) будут использоваться конечные разности интерполируемой функции $f(x)$ (см. (1.19) в § 1.4).

Имеем:

$$\delta_1 := -\frac{1}{4}, \quad \lambda_1 := \frac{3}{4h^2} \Delta^2 f_0;$$

при $k = 3, 4, \dots, n$:

$$\delta_{k-1} = -\frac{1}{4 + \delta_{k-2}}, \quad \lambda_{k-1} = \frac{\frac{3}{h^2} \Delta^2 f_{k-2} - \lambda_{k-2}}{4 + \delta_{k-2}};$$

$$c_n := 0;$$

при $k = n, n-1, \dots, 2$:

$$c_{k-1} = \delta_{k-1} c_k + \lambda_{k-1};$$

при $k = 1, 2, \dots, n$ (с учетом $c_0 := 0$):

$$b_k = \frac{\Delta f_{k-1}}{h} + \frac{h}{3} (2c_k + c_{k-1}), \quad d_k = \frac{c_k - c_{k-1}}{3h}.$$

В результате при $x \in [x_{k-1}, x_k]$ значение $f(x)$ может быть заменено значением

$$g_k(x) = f_k + b_k(x - x_k) + c_k(x - x_k)^2 + d_k(x - x_k)^3$$

с найденными значениями коэффициентов b_k, c_k, d_k и погрешностью $O(h^4)$ в соответствии с оценкой (4.28).

Замечание 4.2. Если известно, что значения f_k приближаемой на $[a, b]$ функции $f(x)$ содержат случайные ошибки, то вместо интерполяционного сплайна применяют *сглаживающий сплайн*. Сглаживающим сплайном, наиболее близким к естественному сплайну третьей степени, является такой кубический n -звенник $S_3(x)$, который минимизирует функционал

$$\Phi_1(u) := \int_a^b (u''(x))^2 dx + \sum_{k=0}^n p_k [u(x_k) - f_k]^2,$$

где $p_k > 0$ — некоторые весовые коэффициенты, от величины которых зависит степень сглаживания (или иначе, допустимое отклонение сглаживающего сплайна от интерполяционного в каждом конкретном узле) [14, 63 и др.]. Процедура построения сглаживающих сплайнов более сложна, и для получения коэффициентов, например, кубического сглаживающего сплайна $S_3(x)$ требуется решать линейные алгебраические системы с пятидиагональными матрицами [63, 71].

§ 4.3. КВАДРАТИЧНЫЙ СПЛАЙН ДЕФЕКТА 1

Следуя [84] (см. также [37, 62 и др.]), будем строить такой интерполяционный сплайн второй степени, узлы которого не совпадают с узлами интерполяции, а перемежаются с ними (за исключением начального и последнего). А именно, пусть аппроксимируемая функция $f(x)$ на системе $n+1$ точек

$$a = x_0 < x_1 < x_2 < \dots < x_n = b$$

— узлов интерполяции — принимает значения $f(x_k) = f_k$ ($k = 0, 1, \dots, n$), и пусть все внутренние узлы $\tilde{x}_k$ **квадратичного сплайна** $p(x)$ берутся точно посередине между соседними узлами интерполяции, т.е.

$$\tilde{x}_k = \frac{1}{2}(x_{k-1} + x_k),$$

а крайними узлами сплайна служат концы отрезка $[a, b]$:

$$\tilde{x}_0 := a (= x_0), \quad \tilde{x}_{n+1} := b (= x_n).$$

Таким образом, на каждом из $n+1$ отрезков $[\tilde{x}_{k-1}, \tilde{x}_k]$, где $k = 1, 2, \dots, n+1$ (рис. 4.5), сплайн $p(x)$ определяется своим **звеном** $p_k(x)$, которое будем задавать квадратичной функцией вида

$$p_k(x) := a_k + b_k(x - \tilde{x}_k) + c_k(x - \tilde{x}_k)^2. \quad (4.29)$$

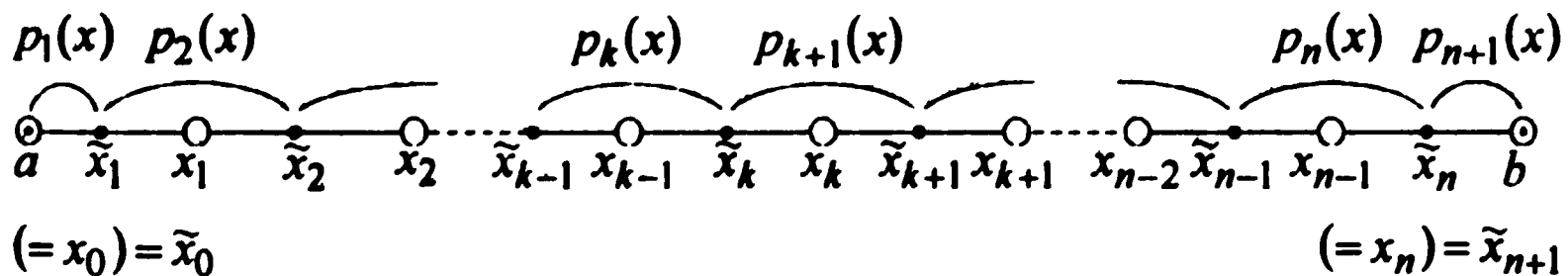


Рис. 4.5. Расположение узлов и звеньев квадратичного сплайна

Чтобы квадратичный сплайн

$$p(x) := \{ p_k(x), \quad x \in [\tilde{x}_{k-1}, \tilde{x}_k] \}_{k=1}^{n+1}$$

был интерполяционным для функции $f(x)$ и имел дефект 1, нужно удовлетворить условиям интерполяции во всех точках согласования функций $f(x)$ и $p(x)$:

$$p_{k+1}(x_k) = f_k \quad \text{при } k = 0, 1, \dots, n, \quad (4.30)$$

и условиям гладкой стыковки звеньев сплайна во всех его внутренних узлах:

$$\left. \begin{aligned} p_{k+1}(\tilde{x}_k) &= p_k(\tilde{x}_k) \\ p'_{k+1}(\tilde{x}_k) &= p'_k(\tilde{x}_k) \end{aligned} \right| \quad \text{при } k = 1, 2, \dots, n. \quad (4.31)$$

В совокупности выписанные условия дают $3n+1$ уравнений относительно $3n+3$ неизвестных коэффициентов a_k, b_k, c_k , фигурирующих в выражениях функций $p_k(x)$ (4.29), составляющих $p(x)$. Оставшиеся свободными две связи можно реализовать по-разному; остановимся на задании краевых условий на вторую производную сплайна^{*}):

$$p''_1(a) = A, \quad p''_{n+1}(b) = B. \quad (4.32)$$

Продифференцировав дважды функцию (4.29), на основе условий (4.30)–(4.32) получаем следующую систему уравнений:

$$\left\{ \begin{aligned} a_{k+1} + b_{k+1}(x_k - \tilde{x}_{k+1}) + c_{k+1}(x_k - \tilde{x}_{k+1})^2 &= f_k \quad (k=0, 1, \dots, n), \\ a_{k+1} + b_{k+1}(\tilde{x}_k - \tilde{x}_{k+1}) + c_{k+1}(\tilde{x}_k - \tilde{x}_{k+1})^2 &= a_k, \\ b_{k+1} + 2c_{k+1}(\tilde{x}_k - \tilde{x}_{k+1}) &= b_k, \\ 2c_1 &= A, \quad 2c_{n+1} = B. \end{aligned} \right\} \quad (k=1, 2, \dots, n) \quad (4.33)$$

Для более удобного манипулирования с этой системой введем шаг интерполяции

$$h_k := x_k - x_{k-1}, \quad (4.34)$$

через который выразим фигурирующие в ней величины:

$$x_0 - \tilde{x}_1 = -\frac{1}{2}h_1, \quad x_k - \tilde{x}_{k+1} = -\frac{1}{2}h_{k+1}, \quad \tilde{x}_k - \tilde{x}_{k+1} = -\frac{1}{2}(h_k + h_{k+1})$$

при $k = 1, 2, \dots, n-1$ и

$$x_n - \tilde{x}_{n+1} = 0, \quad \tilde{x}_n - \tilde{x}_{n+1} = -\frac{1}{2}h_n \quad \text{при } k = n.$$

^{*} В [84] показывается, что задание обеих дополнительных связей на одном конце промежутка $[a, b]$ приводит к численно неустойчивому процессу сплайн-интерполяции.

С обозначениями (4.34) система (4.33) приобретает вид

$$\left\{ \begin{array}{l} a_{k+1} - \frac{1}{2}h_{k+1}b_{k+1} + \frac{1}{4}h_{k+1}^2c_{k+1} = f_k \quad (k = 0, 1, \dots, n-1), \\ a_{n+1} = f_n, \\ a_{k+1} - \frac{1}{2}(h_k + h_{k+1})b_{k+1} + \frac{1}{4}(h_k + h_{k+1})^2c_{k+1} = a_k \\ \quad (k = 1, 2, \dots, n-1), \\ a_{n+1} - \frac{1}{2}h_nb_{n+1} + \frac{1}{4}h_n^2c_{n+1} = a_n, \\ b_{k+1} - (h_k + h_{k+1})c_{k+1} = b_k \quad (k = 1, 2, \dots, n-1), \\ b_{n+1} - h_nc_{n+1} = b_n, \\ 2c_1 = A, \quad 2c_{n+1} = B. \end{array} \right. \quad (4.35)$$

Результат преобразований этой системы будет приведен позже (см. формулы (4.52)–(4.55)), а сейчас рассмотрим упрощенный случай, когда параболическая сплайн-интерполяция совершается по системе равноотстоящих узлов.

Будем считать, что шаг интерполяции h_k постоянен и равен $2h$, т.е. пусть в (4.34)

$$h_k = 2h \quad \forall k \in \{1, 2, \dots, n\}.$$

Тогда коэффициенты a_k , b_k , c_k звеньев сплайна (4.29) могут быть найдены из следующего частного случая системы (4.35):

$$\left\{ \begin{array}{l} a_{k+1} - hb_{k+1} + h^2c_{k+1} = f_k \quad (k = 0, 1, \dots, n-1), \end{array} \right. \quad (4.36)$$

$$a_{n+1} = f_n, \quad (4.37)$$

$$a_{k+1} - 2hb_{k+1} + 4h^2c_{k+1} = a_k \quad (k = 1, 2, \dots, n-1), \quad (4.38)$$

$$\left\{ \begin{array}{l} a_{n+1} - hb_{n+1} + h^2c_{n+1} = a_n, \end{array} \right. \quad (4.39)$$

$$b_{k+1} - 4hc_{k+1} = b_k \quad (k = 1, 2, \dots, n-1), \quad (4.40)$$

$$b_{n+1} - 2hc_{n+1} = b_n, \quad (4.41)$$

$$\left\{ \begin{array}{l} 2c_1 = A, \quad 2c_{n+1} = B. \end{array} \right. \quad (4.42)$$

Как и в предыдущем параграфе, будем исключать из уравнений системы неизвестные a_k , b_k и сводить всё к связям между неизвестными c_k .

Так как из (4.36)

$$a_{k+1} = hb_{k+1} - h^2 c_{k+1} + f_k \quad \text{при } k = 0, 1, \dots, n-1,$$

и следовательно,

$$a_k = hb_k - h^2 c_k + f_{k-1} \quad \text{при } k = 1, 2, \dots, n, \quad (4.43)$$

то, воспользовавшись этим в равенстве (4.38), имеем

$$hb_{k+1} - h^2 c_{k+1} + f_k - hb_k + h^2 c_k - f_{k-1} - 2hb_{k+1} + 4h^2 c_{k+1} = 0,$$

т.е. при любых $k \in \{1, 2, \dots, n-1\}$ справедливо равенство

$$b_k + b_{k+1} = hc_k + 3hc_{k+1} + \frac{\Delta f_{k-1}}{h} \quad (4.44)$$

(напомним, что $\Delta f_{k-1} := f_k - f_{k-1}$ — конечная разность первого порядка функции $f(x)$). Рассматривая (4.44) совместно с равенством (4.40), получаем

$$2b_k = hc_k - hc_{k+1} + \frac{\Delta f_{k-1}}{h} \quad \text{при } k = 1, 2, \dots, n-1, \quad (4.45)$$

и значит,

$$2b_{k+1} = hc_{k+1} - hc_{k+2} + \frac{\Delta f_k}{h} \quad \text{при } k = 0, 1, \dots, n-2. \quad (4.46)$$

Подставим (4.45) и (4.46) в равенство (4.44) (предварительно умноженное на 2):

$$hc_k - hc_{k+1} + \frac{\Delta f_{k-1}}{h} + hc_{k+1} - hc_{k+2} + \frac{\Delta f_k}{h} = 2hc_k + 6hc_{k+1} + \frac{2\Delta f_{k-1}}{h}.$$

Отсюда после приведения подобных членов, деления на h и введения конечной разности второго порядка $\Delta^2 f_{k-1} := \Delta f_k - \Delta f_{k-1}$ приходим к трехточечному рекуррентному равенству

$$c_k + 6c_{k+1} + c_{k+2} = \frac{\Delta^2 f_{k-1}}{h^2}, \quad (4.47)$$

связывающему неизвестные величины c_k при $k = 1, 2, \dots, n-2$.

При последнем из допустимых в (4.47) значений $k = n-2$ получается уравнение, содержащее неизвестные c_{n-2} , c_{n-1} и c_n . Чтобы вывести уравнение, связывающее c_{n-1} , c_n и c_{n+1} , вос-

пользуемся нетронутыми еще соотношениями (4.37), (4.39) и (4.41).

Сначала подставим $a_{n+1} = f_n$ и $a_n = hb_n - h^2c_n + f_{n-1}$ (см. (4.36) при $k = n-1$) в (4.39), в результате чего получим

$$b_{n+1} + b_n = hc_n + hc_{n+1} + \frac{\Delta f_{n-1}}{h}.$$

Рассматривая это равенство совместно с (4.41), т.е. с $b_{n+1} - b_n = 2hc_{n+1}$, находим выражение

$$2b_n = hc_n - hc_{n+1} + \frac{\Delta f_{n-1}}{h}. \quad (4.48)$$

Теперь обратимся к равенству (4.45), согласно которому при $k = n-1$ имеем

$$2b_{n-1} = hc_{n-1} - hc_n + \frac{\Delta f_{n-2}}{h}. \quad (4.49)$$

Но, в силу (4.44) при $k = n-1$, должно быть

$$2b_n + 2b_{n-1} = 2hc_{n-1} + 6hc_n + \frac{2\Delta f_{n-2}}{h}.$$

Складывая равенства (4.48) и (4.49) и сравнивая полученное с последним равенством, после упрощений приходим к равенству

$$c_{n-1} + 6c_n + c_{n+1} = \frac{\Delta^2 f_{n-2}}{h^2}, \quad (4.50)$$

имеющему абсолютно ту же структуру, что и (4.47); иначе говоря, (4.50) — это (4.47) при $k = n-1$.

Итак, собирая воедино формулы, нужные для цельного описания процедуры построения интерполяционного квадратичного сплайна

$$p(x) := \{ p_k(x), \quad x \in [\tilde{x}_{k-1}, \tilde{x}_k] \}_{k=1}^{n+1}$$

дефекта 1 по системе равноотстоящих с шагом $2h$ узлов интерполяции x_k и узлов сплайна $\tilde{x}_k = x_{k-1} + h$, имеем следующее.

Решая трехдиагональную систему

$$\left\{ \begin{array}{l} 2c_1 = A, \\ 6c_2 + c_3 = \frac{\Delta^2 f_0}{h^2} - \frac{A}{2}, \\ c_2 + 6c_3 + c_4 = \frac{\Delta^2 f_1}{h^2}, \\ \dots\dots\dots \\ c_{n-2} + 6c_{n-1} + c_n = \frac{\Delta^2 f_{n-3}}{h^2}, \\ c_{n-1} + 6c_n = \frac{\Delta^2 f_{n-2}}{h^2} - \frac{B}{2}, \\ 2c_{n+1} = B \end{array} \right. \quad (4.51)$$

(например, прогонкой, которая здесь заведомо корректна и устойчива), находим коэффициенты c_k ($k = 1, 2, \dots, n+1$) звеньев сплайна $p_k(x)$ (4.29); обоснованием этой системы служат формулы (4.42), (4.47), (4.50). По известным c_k далее вычисляем

$$b_k = \frac{h}{2}(c_k - c_{k+1}) + \frac{1}{2} \frac{\Delta f_{k-1}}{h} \quad \text{при} \quad k = 1, 2, \dots, n$$

в соответствии с (4.45) и (4.48); после этого находим

$$b_{n+1} = b_n + Bh$$

на основе (4.41) с учетом (4.42). Последними вычисляются коэффициенты a_k (см. (4.43) и (4.37)):

$$\begin{aligned} a_k &= h(b_k - hc_k) + f_{k-1} \quad \text{при } k = 1, 2, \dots, n, \\ a_{n+1} &= f_n. \end{aligned}$$

Замечание 4.3. Если сравнить систему (4.51) с аналогичной системой, получающейся из (4.23) при $h_k = h$ в случае построения кубического сплайна по равноотстоящим узлам, можно заметить, что в системе (4.51) более ярко выражено диагональное преобладание; это позволяет надеяться на лучшую численную устойчивость процедуры нахождения коэффициентов квадратичного сплайна.

Вывод формул для получения коэффициентов c_k , b_k , a_k в общем случае переменного шага интерполяции h_k из совокупности условий (4.35) не вызывает затруднений и производится

аналогично продемонстрированному выше выводу. Итогом является следующая ленточная система (где $f(x_{k-1}; x_k; x_{k+1})$ — разделенные разности второго порядка, построенные по значениям f_{k-1}, f_k, f_{k+1}):

$$\left\{ \begin{array}{l} 2c_1 = A, \\ 3c_2 + \frac{h_2}{h_1 + h_2} c_3 = 4f(x_0; x_1; x_2) - \frac{Ah_1}{2h_1 + 2h_2}, \\ \frac{h_2}{h_2 + h_3} c_2 + 3c_3 + \frac{h_3}{h_2 + h_3} c_4 = 4f(x_1; x_2; x_3), \\ \dots\dots\dots \\ \frac{h_{n-2}}{h_{n-2} + h_{n-1}} c_{n-2} + 3c_{n-1} + \frac{h_{n-1}}{h_{n-2} + h_{n-1}} c_n = 4f(x_{n-3}; x_{n-2}; x_{n-1}), \\ \frac{h_{n-1}}{h_{n-1} + h_n} c_{n-1} + 3c_n = 4f(x_{n-2}; x_{n-1}; x_n) - \frac{Bh_n}{2h_{n-1} + 2h_n}, \\ 2c_{n+1} = B. \end{array} \right. \quad (4.52)$$

Решив ее относительно величин c_k , служащих коэффициентами при вторых степенях звеньев сплайна $p_k(x)$, далее вычисляем коэффициенты b_k при первых степенях, полагая $k = 1, 2, \dots, n$ в равенстве

$$b_k = \frac{1}{4}h_k(c_k - c_{k+1}) + f(x_{k-1}; x_k), \quad (4.53)$$

а затем

$$b_{n+1} = b_n + \frac{1}{2} B h_n, \quad (4.54)$$

и, наконец, свободные члены a_k квадратных трехчленов $p_k(x)$, составляющих сплайн $p(x)$, находим с помощью равенств

$$a_k = \frac{1}{2}h_k \left(b_k - \frac{1}{2}h_k c_k \right) + f_{k-1}, \quad (4.55)$$

где $k = 1, 2, \dots, n$ и $a_{n+1} = f_n$.

Проведенные выкладки и их результаты можно рассматривать как доказательство существования и единственности квадратичного сплайна дефекта 1, интерполирующего на отрезке $[a, b]$ заданную на нем своими значениями f_k функцию $f(x)$ при заданных краевых условиях (4.32). Ряд утверждений о

сходимости таких сплайнов к функции $f(x)$ на $[a, b]$ с тем или иным порядком относительно $h \rightarrow 0$ в равномерном и $\Delta := \max(x_k - x_{k-1}) \rightarrow 0$ в неравномерном случаях, в зависимости от требуемой гладкости $f(x)$, можно найти в [84], причем в этих утверждениях даются оценки близости в точках $x \in [a, b]$ не только $f(x)$ и $p(x)$, но и их производных^{*}. В простейшем виде утверждения подобного рода качественно отражает следующая теорема.

Теорема 4.3. Пусть функция $f(x)$ дважды непрерывно дифференцируема на отрезке $[a, b]$ и пусть $p(x)$ — квадратичный сплайн, определяемый условиями (4.30)–(4.32). Тогда при каждом фиксированном n найдутся такие положительные постоянные c_0, c_1, c_2 , что при любом $x \in [a, b]$

$$\begin{aligned} |f(x) - p(x)| &\leq c_0 \Delta^2, \\ |f^{(i)}(x) - p^{(i)}(x)| &\leq c_i \Delta^{2-i} \quad (i = 1, 2), \end{aligned}$$

где $\Delta := \max_{k \in \{1, \dots, n\}} (x_k - x_{k-1})$.

При этом теорема сохраняет силу и в случаях, когда краевые условия (4.32) на вторую производную $p(x)$ подменяются соответствующими условиями на первую производную

$$p'_1(a) = A, \quad p'_{n+1}(b) = B. \quad (4.56)$$

или периодическими краевыми условиями^{**})

$$p_1^{(i)}(a) = p_{n+1}^{(i)}(b), \quad \text{где } i = 1, 2.$$

Замечание 4.4. При построении интерполяционного квадратичного сплайна дефекта 1, соответствующего требованиям (4.30), (4.31) и (4.56), более удобным может оказаться подход, при котором вычисление коэффициентов a_k, b_k, c_k сводится к применению метода прогонки для нахождения значений b_k (а не c_k) из аналогичной (4.52) системы с последующей подстановкой их в формулы для получения a_k и c_k .

^{*} Отметим кстати, что существуют оценки близости производных функции $f(x)$ и кубического сплайна $g(x)$, подобные утверждаемым теоремой 4.3 для квадратичных сплайнов $p(x)$.

^{**} В последнем случае точка a не должна быть узлом сплайна.

§ 4.4. БАЗИСНЫЕ СПЛАЙНЫ

Без сомнений, можно построить ряд других сплайнов, подобных рассмотренным в предыдущих параграфах кубическим и квадратичным сплайнам, которые обладали бы теми или иными заданными свойствами, связанными со свойствами интерполируемых ими на отрезке $[a, b]$ функций $f(x)$. Среди таких сплайнов более высоких степеней стоит особо отметить, пожалуй, лишь сплайны пятой степени дефекта 3, играющие важную роль в приложениях. Существует достаточно общая и глубокая теория сплайнов, изучающая их свойства, представления, сходимость и т. п. [1, 47, 62, 84 и др.]. Имеется много разработок в области практических применений сплайнов [14, 37, 43, 71 и др.].

Остановимся на нескольких специфических конструкциях кусочно-полиномиальных функций, тоже именуемых сплайнами (с некоторыми уточняющими их особенности прилагательными) и определенным образом связанные с обычными сплайнами.

Определение 4.3 [84]. *Базисным сплайном или, короче, B -сплайном степени $m-1$ (дефекта 1) относительно узлов $x_k < x_{k+1} < \dots < x_{k+m}$ называется функция*

$$B_{m-1,k}(x) = B_{m-1}(x_k; x_{k+1}; \dots; x_{k+m}; x) := \\ = m \sum_{i=k}^{k+m} \frac{(x_i - x)_+^{m-1}}{\Pi'_{m,k}(x_i)}, \quad (4.57)$$

где

$$(x_i - x)_+^{m-1} := [\max\{0; (x_i - x)\}]^{m-1}, \quad (4.58)$$

$$\Pi_{m,k}(x) := (x - x_k)(x - x_{k+1}) \dots (x - x_{k+m}). \quad (4.59)$$

Посмотрим, что собой представляют B -сплайны от нулевой до третьей степени в более простом, но и более употребительном случае равноотстоящих узлов B -сплайна, т.е. когда

$$x_{k+i} = x_k + ih, \quad \text{где } i = 0, 1, \dots, m.$$

Сопоставляя обозначение (4.59) с встречавшимся ранее аналогичным обозначением (1.10), без труда «расшифровываем» знаменатели дробей в (4.57): $\Pi'_{m,k}(x_i)$ — это произведение разностей между узлом x_i и всеми остальными узлами от x_k до x_{k+m} .

Числители этих дробей в соответствии с (4.58) есть

$$(x_i - x)_+^{m-1} := \begin{cases} (x_i - x)^{m-1} & \text{при } x \leq x_i, \\ 0 & \text{при } x \geq x_i. \end{cases} \quad (4.60)$$

Дальнейшую детализацию выражений *B*-сплайнов (4.57) будем производить, придавая параметру m значения 1, 2, 3, 4.

1) Пусть $m = 1$. Тогда из (4.57) имеем (считая, что $0^0 := 0$):

$$\begin{aligned} B_{0,k}(x) = B_0(x_k; x_{k+1}; x) &= \frac{(x_k - x)_+^0}{x_k - x_{k+1}} + \frac{(x_{k+1} - x)_+^0}{x_{k+1} - x_k} = \\ &= \begin{cases} -\frac{1}{h} + \frac{1}{h} = 0 & \text{при } x < x_k, \\ 0 + \frac{1}{h} = \frac{1}{h} & \text{при } x_k \leq x < x_{k+1}, \\ 0 + 0 = 0 & \text{при } x \geq x_{k+1}. \end{cases} \end{aligned} \quad (4.61)$$

Таким образом, *B*-сплайн нулевой степени представляет собой функцию-«ступеньку» (см. рис. 4.6)*).

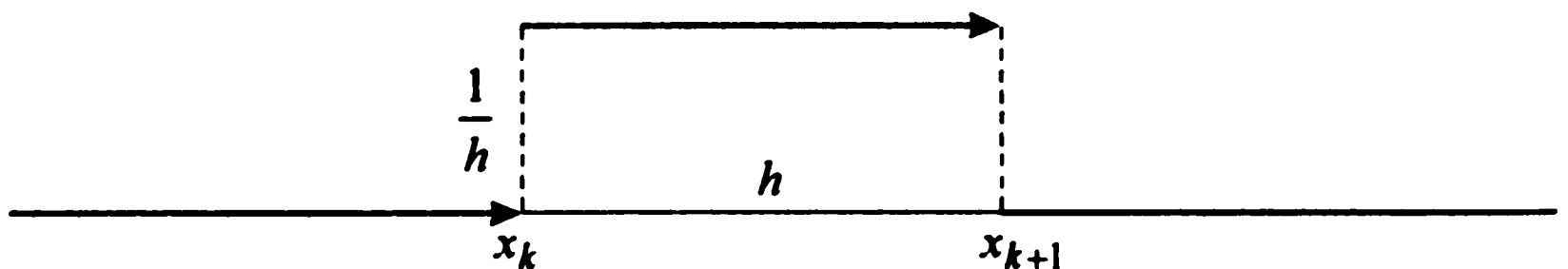


Рис. 4.6. *B*-сплайн нулевой степени

2) При $m = 2$ из (4.57) следует

$$B_{1,k}(x) = B_1(x_k; x_{k+1}; x_{k+2}; x) = 2 \left[\frac{(x_k - x)_+^1}{(x_k - x_{k+1})(x_k - x_{k+2})} + \right.$$

* Это единственный из базисных сплайнов (4.57), изменяющий свои значения скачком. К какому из узлов x_k или x_{k-1} отнести ненулевое значение $\frac{1}{h}$ — вопрос договоренности. То же можно сказать об $(m-1)$ -х производных *B*-сплайнов $(m-1)$ -й степени.

$$\begin{aligned}
& + \frac{(x_{k+1} - x)_+^1}{(x_{k+1} - x_k)(x_{k+1} - x_{k+2})} + \frac{(x_{k+2} - x)_+^1}{(x_{k+2} - x_k)(x_{k+2} - x_{k+1})} \Bigg] = \\
& = \frac{1}{h^2} [(x_k - x)_+ - 2(x_{k+1} - x)_+ + (x_{k+2} - x)_+].
\end{aligned}$$

В соответствии с (4.60) последнее можно представить в виде

$$\begin{cases} \frac{1}{h^2}(x_k - x - 2x_{k+1} + 2x + x_{k+2} - x), & \text{если } x \leq x_k, \\ \frac{1}{h^2}(0 - 2x_{k+1} + 2x + x_{k+2} - x), & \text{если } x_k \leq x \leq x_{k+1}, \\ \frac{1}{h^2}(0 - 0 + x_{k+2} - x), & \text{если } x_{k+1} \leq x \leq x_{k+2}, \\ 0, & \text{если } x \geq x_{k+2}. \end{cases}$$

Простыми упрощениями получаем окончательное выражение **линейного В-сплайна**:

$$B_{1,k}(x) = B_1(x_k; x_{k+1}; x_{k+2}; x) =$$

$$= \begin{cases} 0 & \text{при } x \leq x_k, \\ \frac{1}{h} + \frac{x - x_{k+1}}{h^2} & \text{при } x_k \leq x \leq x_{k+1}, \\ \frac{1}{h} - \frac{x - x_{k+1}}{h^2} & \text{при } x_{k+1} \leq x \leq x_{k+2}, \\ 0 & \text{при } x \geq x_{k+2}. \end{cases} \quad (4.62)$$

Нетрудно представить геометрическое изображение линейного В-сплайна (рис. 4.7) — это так называемая функция-«крышка» (по-другому, функция-«шапочка» [16]).

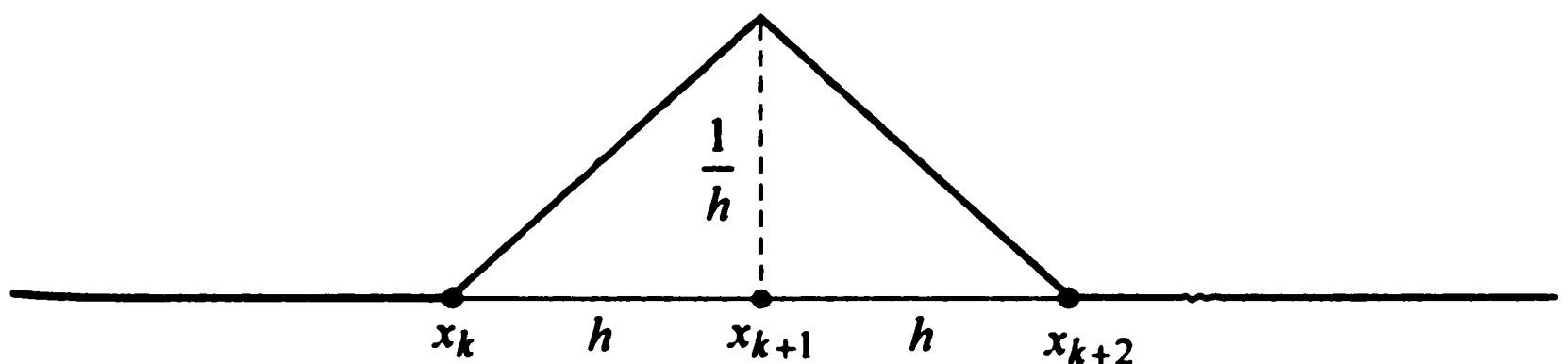


Рис. 4.7. Линейный В-сплайн

3) Полагая в (4.57) $m = 3$ и выполняя несложные преобразования, выясняем, как выглядит **квадратичный В-сплайн**:

$$B_{2,k}(x) = B_2(x_k; x_{k+1}; x_{k+2}; x_{k+3}; x) =$$

$$= \frac{1}{h^3} \left[-\frac{1}{2}(x_k - x)_+^2 + \frac{3}{2}(x_{k+1} - x)_+^2 - \frac{3}{2}(x_{k+2} - x)_+^2 + \frac{1}{2}(x_{k+3} - x)_+^2 \right] =$$

$$= \begin{cases} 0 & \text{при } x \leq x_k, \\ \frac{1}{2h} - \frac{1}{h^2}(x_{k+1} - x) + \frac{1}{2h^3}(x_{k+1} - x)^2 & \text{при } x_k \leq x \leq x_{k+1}, \\ \frac{1}{2h} + \frac{1}{h^2}(x_{k+2} - x) + \frac{1}{h^3}(x_{k+2} - x)^2 & \text{при } x_{k+1} \leq x \leq x_{k+2}, \\ \frac{1}{2h^3}(x_{k+3} - x)^2 & \text{при } x_{k+2} \leq x \leq x_{k+3}, \\ 0 & \text{при } x \geq x_{k+3}. \end{cases} \quad (4.63)$$

График такого сплайна изображен на рис. 4.8. Для его построения предварительно находим

$$B'_{2,k}(x) = \frac{1}{h^3} \begin{cases} h - (x_{k+1} - x), & x_k \leq x \leq x_{k+1}, \\ -h + 2(x_{k+2} - x), & x_{k+1} \leq x \leq x_{k+2}, \\ -(x_{k+3} - x), & x_{k+2} \leq x \leq x_{k+3}, \end{cases}$$

откуда следует, что производная функции $B_{2,k}(x)$ непрерывна во

всех ее узлах и что $\max B_{2,k}(x) = \frac{3}{4h}$ в точке $x = x_{k+1} + \frac{h}{2} = x_{k+2} - \frac{h}{2}$. Из выражения второй производной

$$B''_{2,k}(x) = \frac{1}{h^3} \begin{cases} 1, & x_k \leq x < x_{k+1}, \\ -2, & x_{k+1} \leq x < x_{k+2}, \\ 1, & x_{k+2} \leq x < x_{k+3} \end{cases}$$

видим, что выпуклость функции $B_{2,k}(x)$ изменяется в узлах скачком.

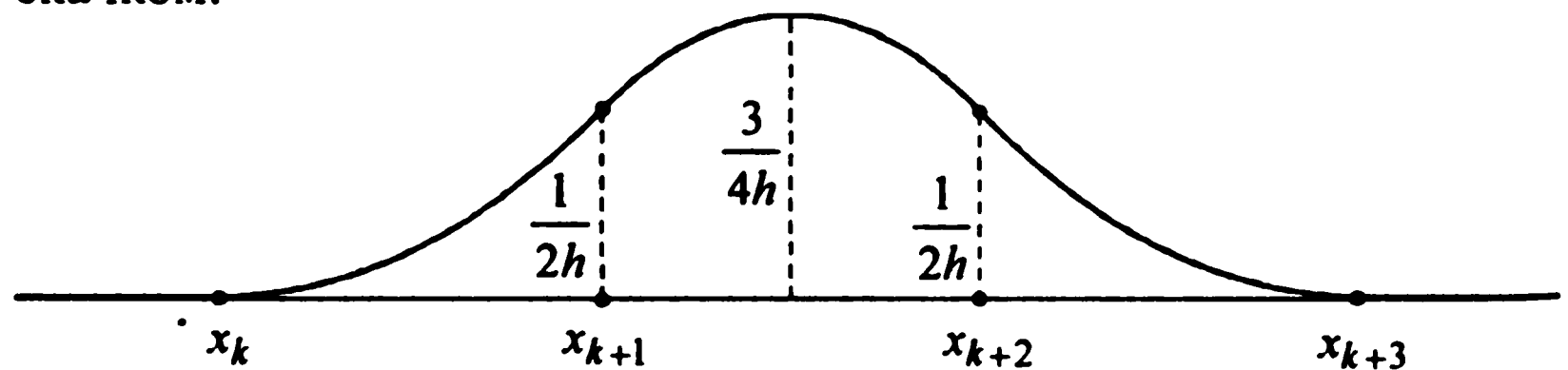


Рис. 4.8. Квадратичный В-сплайн

4) При $m = 4$ аналогичным образом нетрудно получить из (4.57) **кубический В-сплайн**

$$B_{3,k}(x) = \frac{1}{6h^4} \left[(x_k - x)_+^3 - 4(x_{k+1} - x)_+^3 + 6(x_{k+2} - x)_+^3 - \right. \\ \left. - 4(x_{k+3} - x)_+^3 + (x_{k+4} - x)_+^3 \right] =$$

$$= \begin{cases} 0, & x \leq x_k, \\ \frac{1}{6h^4}(x - x_k)^3, & x_k \leq x \leq x_{k+1}, \\ \frac{1}{6h} + \frac{1}{2h^2}(x - x_{k+1}) + \frac{1}{2h^3}(x - x_{k+1})^2 - \\ - \frac{1}{2h^4}(x - x_{k+1})^3, & x_{k+1} \leq x \leq x_{k+2}, \\ \frac{1}{6h} + \frac{1}{2h^2}(x_{k+3} - x) + \frac{1}{2h^3}(x_{k+3} - x)^2 - \\ - \frac{1}{2h^4}(x_{k+3} - x)^3, & x_{k+2} \leq x \leq x_{k+3}, \\ \frac{1}{6h^4}(x_{k+4} - x)^3, & x_{k+3} \leq x \leq x_{k+4}, \\ 0, & x \geq x_{k+4}. \end{cases}$$

Так же легко убедиться, что базисный сплайн третьей степени имеет непрерывными не только первые производные, как предыдущий, но и вторые. График этого сплайна представлен на рис. 4.9^{*)}.

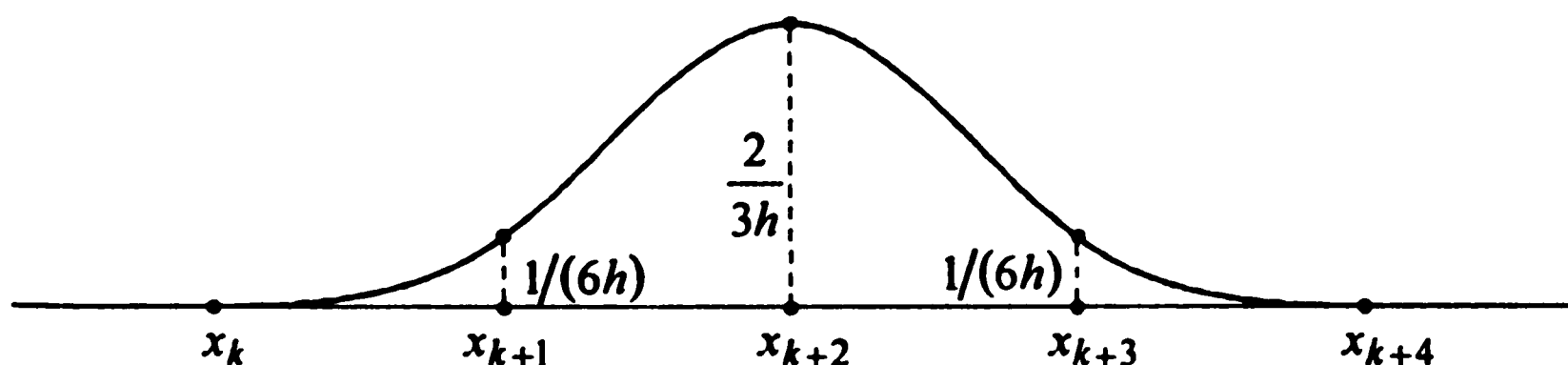


Рис. 4.9. Кубический В-сплайн

Являясь линейно независимыми функциями, В-сплайны определенной степени образуют базис в соответствующих

^{*)} Выведенный здесь из общей формулы кубический В-сплайн отличается от аналогичного сплайна, например, из [73] множителем $\frac{2}{3h}$.

функциональных пространствах, что оправдывает их название и говорит о возможности представления через B -сплайны других функций этих пространств. Интуитивно ясно, что любая кусочно-постоянная функция на отрезке, составленном из равных промежутков длины h , может быть единственным образом представлена линейной комбинацией B -сплайнов нулевой степени (4.61), любая кусочно-линейная — B -сплайнов первой степени (4.62), и т.д. Эти факты могут быть четко сформулированы и строго обоснованы. Например, для изученного в § 4.2 естественного сплайна справедливо следующее утверждение [73].

Теорема 4.4. Пусть $g(x)$ — кубический сплайн (4.8) дефекта 1, построенный по системе равноотстоящих узлов $x_k = x_0 + kh$ ($k = 0, 1, \dots, n$). Тогда найдутся такие постоянные $\alpha_0, \alpha_1, \dots, \alpha_{n+2}$, что

$$g(x) = \sum_{i=0}^{n+2} \alpha_i B_{3,i-3}(x). \quad (4.64)$$

Заметим, что для построения функций $B_{3,-3}(x)$, $B_{3,-2}(x)$, $B_{3,-1}(x)$ и $B_{3,n-3}(x)$, $B_{3,n-2}(x)$, $B_{3,n-1}(x)$ в (4.64) требуется введение вспомогательных точек сетки (за пределами отрезка $[a, b]$) x_{-3} , x_{-2} , x_{-1} и x_{n+1} , x_{n+2} , x_{n+3} соответственно.

Приближение функций линейными комбинациями базисных сплайнов позволяет запоминать лишь коэффициенты α_i этих комбинаций. Вычисление же их, в свою очередь, может осуществляться достаточно эффективно, в силу очевидной ленточной структуры матриц линейных систем, к которым сводится нахождение α_i приравниванием левых и правых частей равенств типа (4.64) в узлах сплайнов $x = x_k$.

B -сплайны играют громадную роль при построении численно-аналитических методов решения дифференциальных и интегральных уравнений. В частности, линейные B -сплайны (4.62) лежат в основе весьма популярного метода конечных элементов (см. гл. 10).

§ 4.5. ЭРМИТОВЫ (ЛОКАЛЬНЫЕ) СПЛАЙНЫ.

Предположим, что у аппроксимируемой функции $y = f(x)$ в точках $x_0, x_1, \dots, x_n \in [a, b]$ известны не только значения $y_0, y_1, \dots, y_n$, но и значения производных до m -й включительно,

т.е. функция $y = f(x)$ задается таблицей

x	y	y'	...	$y^{(m)}$
x_0	y_0	y'_0	...	$y_0^{(m)}$
x_1	y_1	y'_1	...	$y_1^{(m)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_n	y_n	y'_n	...	$y_n^{(m)}$

(4.65)

Как известно (см. § 1.8), по этой информации об $f(x)$ можно построить единственный интерполяционный многочлен Эрмита степени $(n+1)(m+1)-1$. Даже при сравнительно небольших значениях n и m эта степень может оказаться неоправданно высокой. Здесь имеется в виду как сложность построения и эффективного использования многочленов Эрмита высоких степеней, так и повышенные требования к гладкости функций, для которых они должны служить приближениями.

Применение в таких случаях обычных интерполяционных сплайнов высоких степеней и малых дефектов дает нужную гладкость интерполируемой функции, т. е. самого сплайна, но не обеспечивает согласования его производных с данными производными функции $f(x)$.

Напрашивается сведение воедино этих двух способов аппроксимации функций, т. е. образование гибрида интерполяционного сплайна и интерполяционного многочлена Эрмита. Такие гибриды называют **эрмитовыми сплайнами**.

Глядя на таблицу (4.65), в соответствии с полученным из § 1.8 представлением об эрмитовой интерполяции, легко вообразить, что по данным в точках x_0 и x_1 значениям функции $f(x)$ и ее m производных можно построить единственный многочлен Эрмита $H_{2m+1}^1(x)$ степени $2m+1$, по данным в точках x_1 и x_2 — многочлен такой же степени $H_{2m+1}^2(x)$, и т. д. Поскольку каждый из этих многочленов — звеньев эрмитова сплайна

$$S_{2m+1}(x) := \left\{ H_{2m+1}^i(x), x \in [x_{i-1}, x_i] \right\}_{i=1}^n \quad (4.66)$$

— строится независимо от остальных, то такие сплайны называются также **локальными сплайнами**.

Свойство локальности эрмитова сплайна дает определенные преимущества при его построении, особенно в случае большого числа n элементарных промежутков $[x_{i-1}, x_i]$, а также позволяет составлять эрмитов сплайн из звеньев разных степеней, если на каких-то участках промежутка интерполяции $[a, b]$ функция $f(x)$ представлена производными одного порядка, а на

других — другого порядка. При этом заметим, что в точках стыковки таких участков допустимо использование односторонних производных.

Посмотрим, что собой представляет эрмитов сплайн, например, третьей степени (эрмитов сплайн первой степени, по сути, определен в § 4.1 формулами (4.2)–(4.3)).

Считая, что на i -м элементарном промежутке функция $y = f(x)$ задается таблицей

x	y	y'
x_{i-1}	y_{i-1}	y'_{i-1}
x_i	y_i	y'_i

полагаем в (4.66) $m = 1$, и i -е звено сплайна $S_3(x)$ ищем в виде^{*)}

$$H_3^i(x) = a_0^i + a_1^i(x - x_{i-1}) + a_2^i(x - x_{i-1})^2 + a_3^i(x - x_{i-1})^3. \quad (4.67)$$

Два уравнения для нахождения коэффициентов многочлена $H_3^i(x)$ получаем из интерполяционных условий $H_3^i(x_{i-1}) = y_{i-1}$, $H_3^i(x_i) = y_i$, а именно:

$$a_0^i = y_{i-1}, \quad (4.68)$$

$$a_0^i + a_1^i h_i + a_2^i h_i^2 + a_3^i h_i^3 = y_i, \quad (4.69)$$

где $h_i := x_i - x_{i-1}$. Далее дифференцируем (4.67):

$$(H_3^i)'(x) = a_1^i + 2a_2^i(x - x_{i-1}) + 3a_3^i(x - x_{i-1})^2,$$

и из условий $(H_3^i)'(x_{i-1}) = y'_{i-1}$, $(H_3^i)'(x_i) = y'_i$ получаем еще два уравнения:

$$a_1^i = y'_{i-1}, \quad (4.70)$$

$$a_1^i + 2a_2^i h_i + 3a_3^i h_i^2 = y'_i. \quad (4.71)$$

Равенства (4.68) и (4.70) уже определяют первые коэффициенты многочлена $H_3^i(x)$. Подставив их в оставшиеся равенства (4.69) и (4.71), легко находим третий и четвертый коэффициенты:

$$a_2^i = \frac{1}{h_i^2} (3y_i - 3y_{i-1} - 2h_i y'_{i-1} - h_i y'_i), \quad (4.72)$$

$$a_3^i = \frac{1}{h_i^3} (2y_{i-1} - 2y_i + h_i y'_{i-1} + h_i y'_i). \quad (4.73)$$

^{*)} Здесь i — определяемый номером узла индекс, а не показатель степени.

Обратим внимание на то, что формула (4.66) определяет эрмитов сплайн заведомо нечетной степени. Для построения эрмитовых сплайнов четных степеней постановка задачи несколько видоизменяется.

Как и в случае рассматриваемых ранее (см. § 4.3) квадратичных интерполяционных сплайнов, будем строить эрмитов сплайн $S_{2m}(x)$ для заданной таблицей (4.65) функции $y = f(x)$, вводя дополнительные узлы z_i между данными узлами интерполяции x_{i-1} и x_i ; для простоты будем считать, что z_i находится точно посередине между x_{i-1} и x_i . Тогда одно звено $G_{2m}^i(x)$ сплайна $S_{2m}(x)$ на элементарном промежутке $[x_{i-1}, x_i]$ составляется из двух многочленов:

$$G_{2m}^i(x) := \begin{cases} p_{2m}^i(x) := a_0^i + a_1^i(x - x_{i-1}) + \dots + a_{2m}^i(x - x_{i-1})^{2m}, \\ \text{если } x \in [x_{i-1}, z_i], \\ q_{2m}^i(x) := b_0^i + b_1^i(x_i - x) + \dots + b_{2m}^i(x_i - x)^{2m}, \\ \text{если } x \in [z_i, x_i]. \end{cases}$$

Для нахождения $2(2m+1)$ коэффициентов многочленов-полузвеньев $p_{2m}^i(x)$ и $q_{2m}^i(x)$ i -го звена $G_{2m}^i(x)$ сплайна $S_{2m}(x)$ (дефекта m) привлекаются $2m+2$ интерполяционных условия

$$(G_{2m}^i)^{(j)}(x_{i-1}) = y_{i-1}^{(j)}, \quad (G_{2m}^i)^{(j)}(x_i) = y_i^{(j)} \quad (j = 0, 1, \dots, m)$$

и $2m$ условий гладкой стыковки полузвеньев в точке z_i :

$$(p_{2m}^i)^{(j)}(z_i) = (q_{2m}^i)^{(j)}(z_i) \quad (j = 0, 1, \dots, 2m-1).$$

Несложно доказать, что такие коэффициенты могут быть однозначно найдены без привлечения информации с других промежутков, что означает возможность построения локальных сплайнов четной степени.

Построим квадратичный эрмитов сплайн, пользуясь теми же данными, которые использовались при построении кубического эрмитова сплайна.

Представляя i -е звено эрмитова сплайна $S_2(x)$ многочленами

$$p_2^i(x) = a_0^i + a_1^i(x - x_{i-1}) + a_2^i(x - x_{i-1})^2, \quad \text{если } x \in [x_{i-1}, z_i], \quad (4.74)$$

и

$$q_2^i(x) = b_0^i + b_1^i(x_i - x) + b_2^i(x_i - x)^2, \quad \text{если } x \in [z_i, x_i], \quad (4.75)$$

дифференцируем их:

$$(p_2^i)'(x) = a_1^i + 2a_2^i(x - x_{i-1}),$$

$$(q_2^i)'(x) = -b_1^i - 2b_2^i(x_i - x).$$

Далее на основе равенств

$$p_2^i(x_{i-1}) = y_{i-1}, \quad (p_2^i)'(x_{i-1}) = y'_{i-1},$$

$$q_2^i(x_i) = y_i, \quad (q_2^i)'(x_i) = y'_i,$$

$$p_2^i(z_i) = q_2^i(z_i), \quad (p_2^i)'(z_i) = (q_2^i)'(z_i),$$

полагая $z_i - x_{i-1} = x_i - z_i = \frac{1}{2}h_i$, устанавливаем связи между известными величинами y_{i-1} , y_i , y'_{i-1} , y'_i и неизвестными коэффициентами a_0^i , a_1^i , a_2^i , b_0^i , b_1^i , b_2^i i -го звена искомого сплайна:

$$\begin{cases} a_0^i = y_{i-1}, \\ a_1^i = y'_{i-1}, \\ b_0^i = y_i, \\ b_1^i = -y'_i, \\ a_0^i + \frac{1}{2}h_i a_1^i + \frac{1}{4}h_i^2 a_2^i = b_0^i + \frac{1}{2}h_i b_1^i + \frac{1}{4}h_i^2 b_2^i, \\ a_1^i + h_i a_2^i = -b_1^i - h_i b_2^i. \end{cases} \quad (4.76)$$

Первые четыре уравнения полученной системы (4.76) представляют собой явные выражения первых четырех коэффициентов многочленов (4.74) и (4.75); подставив их в последние два уравнения, находим оставшиеся два коэффициента:

$$a_2^i = \frac{2}{h_i^2}(y_i - y_{i-1}) - \frac{1}{2h_i}(y'_i + 3y'_{i-1}), \quad (4.77)$$

$$b_2^i = \frac{2}{h_i^2}(y_{i-1} - y_i) + \frac{1}{2h_i}(3y'_i + y'_{i-1}). \quad (4.78)$$

Пример 4.3. Рассмотрим приближение кубическим и квадратичным эрмитовыми сплайнами функции $y = \sin \pi x$ на отрезке $[0, 1]$, если информация о ней представлена следующей таблицей:

x	y	y'
0	0	π
$\frac{1}{2}$	1	0
1	0	$-\pi$

Полагая $i = 1$ и $i = 2$, по формулам (4.68), (4.70), (4.72) и (4.73) находим коэффициенты кубического эрмитова сплайна, подстановка которых в (4.67) дает

$$S_3(x) = \begin{cases} H_3^1(x) = \pi x + (12 - 4\pi)x^2 + (4\pi - 16)x^3 & \text{при } x \in \left[0, \frac{1}{2}\right], \\ H_3^2(x) = 1 + (2\pi - 12)\left(x - \frac{1}{2}\right)^2 + (16 - 4\pi)\left(x - \frac{1}{2}\right)^3 & \text{при } x \in \left[\frac{1}{2}, 1\right]. \end{cases}$$

Аналогично, с помощью вычисления коэффициентов по формулам (4.76)–(4.78) с последующей их подстановкой в (4.74)–(4.75), приходим к квадратичному эрмитову сплайну

$$S_2(x) = \begin{cases} p_2^1(x) = \pi x + (8 - 3\pi)x^2 & \text{при } x \in \left[0, \frac{1}{4}\right], \\ q_2^1(x) = 1 + (\pi - 8)\left(\frac{1}{2} - x\right)^2 & \text{при } x \in \left[\frac{1}{4}, \frac{1}{2}\right], \\ p_2^2(x) = 1 + (\pi - 8)\left(x - \frac{1}{2}\right)^2 & \text{при } x \in \left[\frac{1}{2}, \frac{3}{4}\right], \\ q_2^2(x) = \pi(1 - x) + (8 - 3\pi)(1 - x)^2 & \text{при } x \in \left[\frac{3}{4}, 1\right]. \end{cases}$$

Точность аппроксимации функции $\sin \pi x$ и ее производной $\pi \cos \pi x$ с помощью построенных эрмитовых сплайнов $S_3(x)$, $S_2(x)$ и их производных покажем в двух контрольных точках следующей таблицей:

x	$\sin \pi x$	$S_3(x)$	$S_2(x)$	$(\sin \pi x)'$	$S_3'(x)$	$S_2'(x)$
$\frac{1}{6}$	0.5	$\frac{2\pi+7}{27} \approx 0.492$	$\frac{2\pi+7}{27} \approx 0.492$	$\frac{\pi\sqrt{3}}{2} \approx 2.72$	$\frac{8}{3} \approx 2.67$	$\frac{8}{3} \approx 2.67$
$\frac{3}{4}$	$\frac{\sqrt{2}}{2} \approx 0.707$	$\frac{\pi+8}{16} \approx 0.696$	$\frac{\pi+8}{16} \approx 0.696$	$\frac{\pi\sqrt{3}}{2} \approx 2.72$	$\frac{\pi}{4} - 3 \approx -2.21$	$\frac{\pi}{2} - 4 \approx -2.43$

Как видно из итоговой таблицы примера 4.3, эрмитова сплайн-интерполяция дает неплохие результаты. Это хорошо согласуется с имеющимися утверждениями о сходимости эрмитовых сплайнов $S_{2m+1}(x)$ и $S_{2m}(x)$ к интерполируемым ими функциям $f(x) \in C_{[a,b]}^m$ с оценкой погрешности $O((\max h_i)^{2m+1})$ и $O((\max h_i)^{2m})$ соответственно (см., например, [47, 84]); для более гладких функций $f(x)$ подобные утверждения устанавливают близость соответствующих производных до m -й степени включительно.

УПРАЖНЕНИЯ

4.1. Выведите формулу осреднения (линейный фильтр) по пяти точкам. Примените ее для сглаживания данных примера 4.2 § 4.1. Результаты сглаживания отобразите графически; проведите визуальное сравнение с результатами осреднения по трем точкам (см. рис.4.3).

4.2. Постройте линейный сплайн $S_1(x)$, интерполирующий функцию $f(x)$, заданную таблицей

x	0	1	2	3
$f(x)$	0	1	1	3

Выразите $S_1(x)$ через линейные B -сплайны (см.(4.62)).

4.3. По данным упр. 4.2 постройте кубический и квадратичный интерполяционные сплайны дефекта 1, полагая $f''(0) = f''(3) = 0$.

4.4. Выведите совокупность формул для построения квадратичного интерполяционного сплайна при краевых условиях типа (4.56) и постоянном шаге $x_k - x_{k-1} = h$, в основу которой легло бы решение трехдиагональной системы относительно коэффициентов b_i звеньев сплайна (4.29) (см. замечание 4.4).

4.5. Постройте квадратичный сплайн дефекта 1, узлы которого $x_0, x_1, \dots, x_n$ совпадали бы с узлами интерполируемой им функции $f(x)$ [13].

4.6. Докажите справедливость формул, фигурирующих в записи кубического B -сплайна $B_{3,k}(x)$.

4.7. Запишите представления и системы уравнений для вычисления коэффициентов этих представлений эрмитовых сплайнов пятой и четвертой степеней.

ГЛАВА 5 || ЧИСЛЕННОЕ ИНТЕГРИРОВАНИЕ

Выводятся и изучаются методы приближенного вычисления определенных интегралов, опирающиеся на суммирование с теми или иными коэффициентами значений подынтегральной функции. А именно, рассматриваются квадратурные формулы прямоугольников, трапеций, Симпсона, Чебышева, Гаусса. Особое внимание уделяется построению эффективных алгоритмов, приводящих с минимальными вычислительными затратами к значению интеграла наперед заданной точности. В их основе — принцип Рунге практического оценивания точности двойным счетом с разными шагами и простые соотношения между некоторыми квадратурными формулами. Затрагивается также проблема вычисления несобственных интегралов.

§ 5.1. ЗАДАЧА ЧИСЛЕННОГО ИНТЕГРИРОВАНИЯ. КВАДРАТУРНЫЕ ФОРМУЛЫ ПРЯМОУГОЛЬНИКОВ

Пусть требуется найти значение I интеграла Римана $\int_a^b f(x)dx$ для некоторой заданной на отрезке $[a, b]$ функции $f(x)$. Хорошо известно, что для функций, допускающих на промежутке $[a, b]$ конечное число точек разрыва первого рода, такое значение существует, единственно и может быть формально получено по определению:

$$I = \lim_{n \rightarrow \infty} \sum_{i=1}^n f(\xi_i)(x_i - x_{i-1}), \quad (5.1)$$

где $\{x_i\}_{i=0}^n$ — произвольная упорядоченная система точек отрезка $[a, b]$ такая, что

$$\max \{x_0 - a, x_i - x_{i-1}, b - x_n\} \rightarrow 0 \quad \text{при} \quad n \rightarrow \infty,$$

а ξ_i — произвольная точка элементарного промежутка $[x_{i-1}, x_i]$.

В математическом анализе обосновывается аналитический способ нахождения значения I с помощью знаменитой формулы Ньютона–Лейбница

$$I = F(b) - F(a), \quad (5.2)$$

где $F(x)$ — некоторая первообразная для данной функции $f(x)$.

К сожалению, применение этого весьма привлекательного подхода к вычислению I наталкивается на несколько серьезных препятствий. Самое главное из них — это несуществование первообразной среди элементарных функций для большинства элементарных функций $f(x)$; например, таким способом не удастся вычислить

$$\int_a^b \frac{\sin x}{x} dx, \quad \int_a^b \frac{dx}{\ln x}, \quad \int_a^b e^{-x^2} dx \quad \text{и т.п.}^*)$$

Если первообразная $F(x)$ для заданной функции $f(x)$ все же найдена, то вычисление двух ее значений $F(a)$ и $F(b)$ может оказаться более трудоемким, чем вычисление существенно большего количества значений $f(x)$.

Поскольку в общем случае значения функций находятся лишь приближенно, использование точной формулы (5.2) приводит к приближенному результату, который может быть более эффективно получен с помощью какой-либо специальной приближенной формулы на основе значений подынтегральной функции $f(x)$. Такие специальные приближенные формулы для вычисления определенных интегралов называют **квадратурными формулами (механическими квадратурами)** или **формулами численного интегрирования**. Первый из этих терминов можно связать с геометрическим смыслом определен-

ного интеграла: вычисление $I := \int_a^b f(x) dx$ при $f(x) \geq 0$ равносильно построению квадрата, равновеликого криволинейной трапеции с основанием $[a, b]$ и «крышей» $f(x)$.

Простые квадратурные формулы можно вывести непосредственно из определения интеграла, т.е. из представления (5.1). Зафиксировав там некоторое $n \geq 1$, будем иметь

$$I \approx \sum_{i=1}^n f(\xi_i)(x_i - x_{i-1}). \quad (5.3)$$

Это приближенное равенство назовем **общей формулой прямоугольников** (площадь криволинейной трапеции приближенно заменяется площадью ступенчатой фигуры, составленной из пря-

^{*)} По главности с этим препятствием к принятию формулы (5.2) может конкурировать разве что возможная дискретность задания подынтегральной функции $f(x)$, довольно характерная для многих реальных приложений определенного интеграла.

моугольников, основаниями которых служат отрезки $[x_{i-1}, x_i]$, а высотами — ординаты $f(\xi_i)$, рис. 5.1).

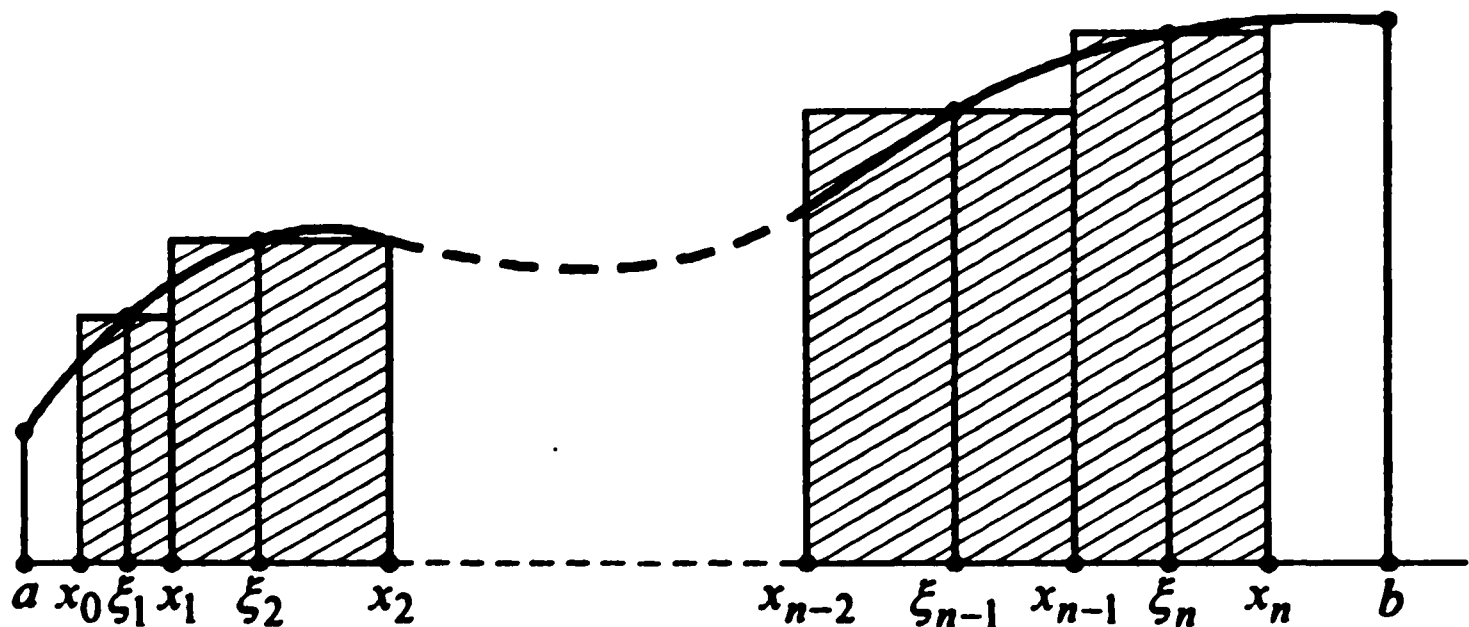


Рис. 5.1. Геометрическая интерпретация общей формулы прямоугольников (5.3)

Чтобы из общей формулы (5.3) получить конструктивное правило приближенного вычисления интеграла, воспользуемся свободой расположения точек x_i , разбивающих промежуток интегрирования $[a, b]$ на элементарные отрезки $[x_{i-1}, x_i]$, и свободой выбора точек ξ_i на этих отрезках.

Условимся в дальнейшем (до § 5.6) пользоваться равномерным разбиением отрезка $[a, b]$ на n частей точками x_i с шагом

$$h = \frac{b-a}{n}, \text{ полагая}$$

$$x_0 = a, \quad x_i = x_{i-1} + h \quad (i = 1, 2, \dots, n-1), \quad x_n = b. \quad (5.4)$$

При таком разбиении формула (5.3) приобретает вид

$$I \approx h \sum_{i=1}^n f(\xi_i), \quad \xi_i \in [x_{i-1}, x_i]. \quad (5.5)$$

Теперь дело за фиксированием точек ξ_i на элементарных отрезках $[x_{i-1}, x_i]$. Рассмотрим три случая.

1) Положим $\xi_i = x_{i-1}$. Тогда из (5.5) получаем

$$I \approx I^{n-} := h \sum_{i=1}^n f(x_{i-1}). \quad (5.6)$$

2) Пусть в (5.5) $\xi_i = x_i$. Тогда имеем

$$I \approx I^{n+} := h \sum_{i=1}^n f(x_i). \quad (5.7)$$

Формулы (5.6) и (5.7) называются **квадратурными формулами левых и правых прямоугольников** соответственно. Совершенно

очевидно (рис. 5.2), что $I^{\Pi-}$ и $I^{\Pi+}$ дают двусторонние приближения к значению I интеграла от монотонной функции.

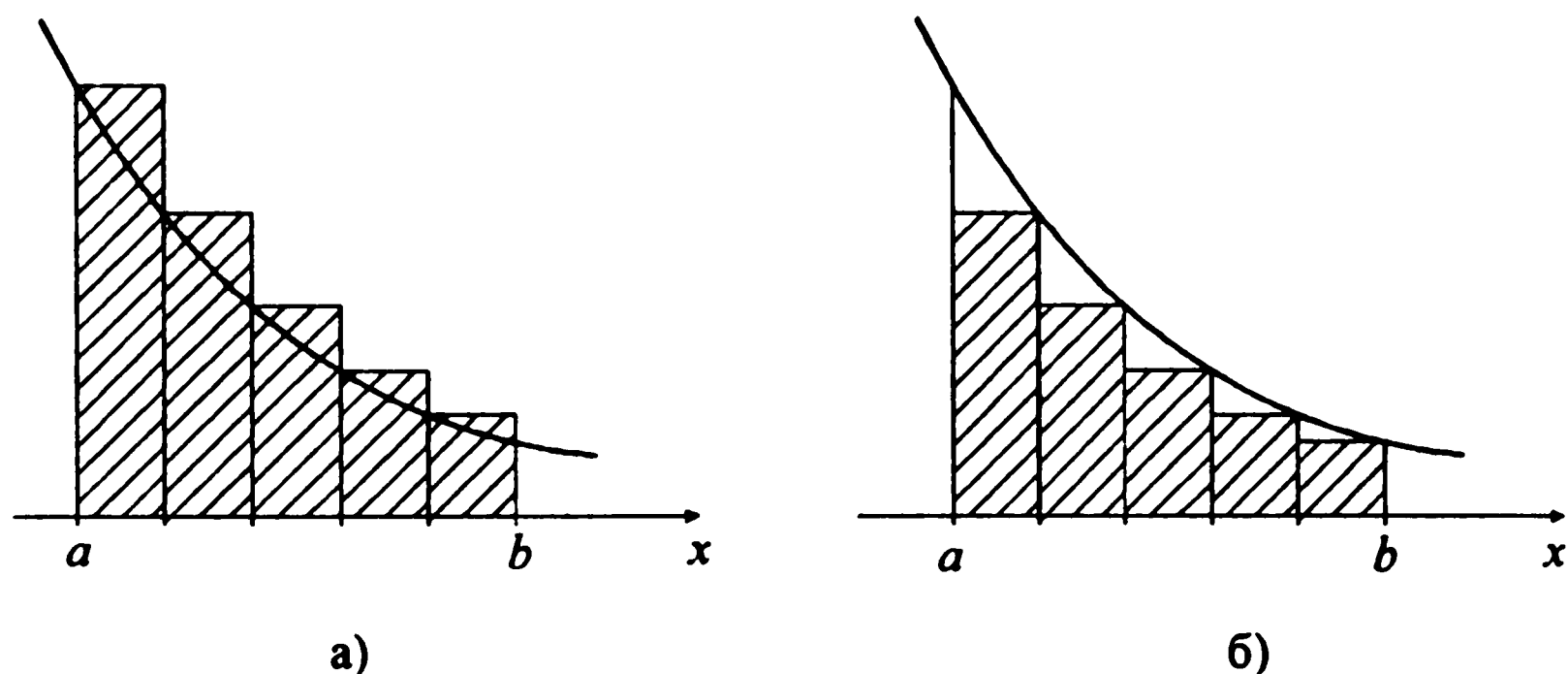


Рис. 5.2. Геометрическое оценивание интеграла от монотонной функции с помощью: а) $I^{\Pi-}$, б) $I^{\Pi+}$.

Можно рассчитывать на бо́льшую точность получения значения интеграла, если взять точку ξ_i посередине между точками x_{i-1} и x_i . Отсюда приходим к следующему случаю.

3) Фиксируем $\xi_i = \frac{1}{2}(x_{i-1} + x_i)$ ($= x_{i-1} + \frac{h}{2} = x_i - \frac{h}{2}$). В результате имеем **квадратурную формулу средних прямоугольников** или, иначе, **формулу средней точки**

$$I \approx I^{\Pi} := h \sum_{i=1}^n f\left(x_{i-1} + \frac{h}{2}\right) = h \sum_{i=1}^n f\left(x_i - \frac{h}{2}\right). \quad (5.8)$$

Учитывая априорно бо́льшую точность формулы (5.8) по сравнению с формулами (5.6) и (5.7) при том же объёме и характере вычислений, эту симметричную формулу будем впредь называть просто **формулой прямоугольников**.

Полученные из определения интеграла квадратурные правила (5.6)–(5.8) не содержат в себе сведений, позволяющих сказать, каким нужно взять число n элементарных промежутков $[x_{i-1}, x_i]$ или, иначе, каким должен быть шаг h разбиения (5.4) отрезка интегрирования $[a, b]$, чтобы гарантированно найти значение I интеграла с наперед заданной точностью ε . Чтобы добыть эти сведения, положим вывод квадратурной формулы прямоугольников на другую основу, предъявив к подынтегральной функции $f(x)$ определенные требования.

Пусть функция $f(x)$ дважды непрерывно дифференцируема.

Рассмотрим сначала вычисление интеграла

$$I_0 := \int_{-\frac{h}{2}}^{\frac{h}{2}} f(x) dx \quad (5.9)$$

при некотором достаточно малом $h > 0$.

По формуле Тейлора можно записать

$$f(x) = f(0) + f'(0)x + \frac{1}{2} f''(\xi)x^2,$$

где x — произвольная, а ξ — некоторая фиксированная точки интервала $\left(-\frac{h}{2}, \frac{h}{2}\right)$. Подставляя это выражение $f(x)$ в (5.9), имеем:

$$I_0 := \int_{-\frac{h}{2}}^{\frac{h}{2}} \left[f(0) + f'(0)x + \frac{1}{2} f''(\xi)x^2 \right] dx = hf(0) + \frac{f''(\xi_0)}{24} h^3, \quad (5.10)$$

где $\xi_0 \in \left(-\frac{h}{2}, \frac{h}{2}\right)$ — некоторая точка, вообще говоря, несовпадающая с точкой ξ (поскольку значение ξ изменяется с измене-

нием h , к $f''(\xi)$ в интеграле $\int_{-\frac{h}{2}}^{\frac{h}{2}} f''(\xi)x^2 dx$ нельзя относиться, как

к постоянной, но, в силу неотрицательности x^2 , можно применить теорему об интегральном среднем, согласно которой, если функция $f_2(x)$ сохраняет определенный знак на отрезке $[a, b]$, то в интервале (a, b) найдется точка c такая, что

$$\int_a^b f_1(x)f_2(x) dx = f_1(c) \int_a^b f_2(x) dx).$$

Слагаемое $hf(0)$ в выражении (5.10) интеграла I_0 очевидно, можно расценивать как приближение к I_0 по формуле прямоугольников в случае всего одного элементарного отрезка.

Второе же слагаемое, т.е. $\frac{f''(\xi_0)}{24} h^3$ есть *остаточный член* (локальная погрешность) простейшей формулы прямоугольников $I_0 \approx hf(0)$.

Вернемся к нахождению значения I интеграла $\int_a^b f(x) dx$ при $f(x) \in C^2[a, b]$.

Выполним разбиение (5.4) отрезка $[a, b]$ на n частей с шагом $h = \frac{b-a}{n}$ и обозначим через $x_{i-\frac{1}{2}}$ середину i -го элементарного отрезка $[x_{i-1}, x_i]$. Тогда каждый отрезок $[x_{i-1}, x_i]$ представляется своим центром симметрии в виде $\left[x_{i-\frac{1}{2}} - \frac{h}{2}, x_{i-\frac{1}{2}} + \frac{h}{2}\right]$, и, следовательно, к каждому интегралу

$$I_i := \int_{x_{i-1}}^{x_i} f(x) dx = \int_{x_{i-\frac{1}{2}} - \frac{h}{2}}^{x_{i-\frac{1}{2}} + \frac{h}{2}} f(x) dx$$

применима формула (5.10) с $x_{i-\frac{1}{2}}$ в роли нуля, т.е.

$$I_i = hf(x_{i-\frac{1}{2}}) + \frac{f''(\xi_i)}{24} h^3,$$

где $i \in \{1, 2, \dots, n\}$, $\xi_i \in (x_{i-1}, x_i)$.

Так как по свойству аддитивности, относящемуся к промежутку интегрирования,

$$\int_a^b f(x) dx = \sum_{i=1}^n \int_{x_{i-1}}^{x_i} f(x) dx, \quad (5.11)$$

то

$$I = \sum_{i=1}^n I_i = \sum_{i=1}^n hf(x_{i-\frac{1}{2}}) + \sum_{i=1}^n \frac{f''(\xi_i)}{24} h^3.$$

В первом слагаемом здесь узнаём выведенную ранее из других соображений формулу прямоугольников I^n (ср. (5.8)), второе же слагаемое характеризует остаточный член формулы прямоугольников

$$r^n(h) := I - I^n = \frac{h^3}{24} \sum_{i=1}^n f''(\xi_i).$$

Согласно обобщенной теореме о среднем значении непрерывной функции, существует точка $\xi_n \in (a, b)$ такая, что

$$\sum_{i=1}^n f''(\xi_i) = \frac{1}{h} \sum_{i=1}^n h f''(\xi_i) = \frac{1}{h} f''(\xi_n) \sum_{i=1}^n h = \frac{b-a}{h} f''(\xi_n).$$

В результате подстановки этого выражения в предыдущее равенство приходим к окончательному виду остаточного члена (*глобальной погрешности*) квадратурной формулы прямоугольников:

$$r^n(h) = \frac{b-a}{24} f''(\xi_n) h^2, \quad \xi_n \in (a, b). \quad (5.12)$$

Как видно из формулы (5.12), при увеличении числа n элементарных отрезков, на которые разбивается промежуток интегрирования $[a, b]$, ошибка численного интегрирования по формуле средней точки (5.8) убывает пропорционально квадрату шага h . Нетрудно убедиться, что погрешность численного интегрирования непрерывно дифференцируемой функции по формулам левых и правых прямоугольников (5.6), (5.7) убывает лишь по линейному закону.

§ 5.2. СЕМЕЙСТВО КВАДРАТУРНЫХ ФОРМУЛ НЬЮТОНА–КОТЕСА

Подстановка в интеграл $\int_a^b f(x) dx$ вместо функции $f(x)$ её интерполяционного многочлена Лагранжа той или иной степени n приводит к семейству квадратурных формул, называемых формулами Ньютона–Котеса^{*)}.

Как было показано в гл.1, функция $f(x) \in C_{[a, b]}^{n+1}$ может быть единственным образом представлена в виде

$$f(x) = L_n(x) + R_n(x),$$

где $L_n(x) = \sum_{i=0}^n \frac{\Pi_{n+1}(x) y_i}{(x - x_i) \Pi'_{n+1}(x_i)}$ — интерполяционный многочлен

Лагранжа, $R_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \Pi_{n+1}(x)$ — остаточный член,

^{*)} Котес Роджер (1682–1716) — английский математик.

$y_i := f(x_i)$, $\Pi_{n+1}(x) := \prod_{i=0}^n (x - x_i)$, $\xi \in (a, b)$. Если система узлов интерполирования $\{x_i\}_{i=0}^n$ совпадает с точками разбиения (5.4) отрезка $[a, b]$ с шагом h , то замена переменной $x = x_0 + qh$ трансформирует многочлен Лагранжа к виду

$$L_n(x_0 + qh) = \sum_{i=0}^n \frac{(-1)^{n-i} y_i}{i!(n-i)!} \cdot \frac{q(q-1)\cdots(q-n)}{q-i}. \quad (5.13)$$

Для того, чтобы использовать такое выражение $L_n(x)$ вместо $f(x)$ в $\int_a^b f(x) dx$, нужно изменить границы интегрирования (значению $x = a$ соответствует значение $q = 0$, а $x = b$ — значение $q = n$) и учесть, что $dx = h dq$. Таким образом, получаем

$$I \approx h \int_0^n \sum_{i=0}^n \frac{(-1)^{n-i} y_i}{i!(n-i)!} \cdot \frac{q(q-1)\cdots(q-n)}{q-i} dq.$$

Это равенство, переписанное в виде

$$I \approx (b-a) \sum_{i=0}^n H_i y_i, \quad (5.14)$$

и есть **квадратурная формула Ньютона–Котеса**, где

$$H_i := \frac{1}{n} \cdot \frac{(-1)^{n-i}}{i!(n-i)!} \int_0^n \frac{q(q-1)\cdots(q-n)}{q-i} dq \quad (5.15)$$

— **коэффициенты Котеса**.

На самом деле, формулы (5.14)–(5.15) определяют семейство квадратурных формул. Параметром этого семейства является число n — степень интерполяционного многочлена, которым заменяется подынтегральная функция.

Рассмотрим несколько простейших частных случаев, соответствующих небольшим значениям $n \in \mathbb{N}$. При этом конкретные формулы будем получать не на основе общих формул (5.14)–(5.15), а используя для этой цели вместо многочлена Лагранжа (5.13) эквивалентный ему (в силу единственности)

первый интерполяционный многочлен Ньютона (1.26):

$$P_n(x_0 + qh) = y_0 + q\Delta y_0 + \frac{q(q-1)}{2!}\Delta^2 y_0 + \dots \\ \dots + \frac{q(q-1)\dots(q-n+1)}{n!}\Delta^n y_0. \quad (5.16)$$

1) Пусть $n=1$, т.е. имеется всего две точки x_0 и $x_1 = x_0 + h$, в которых известны значения функции ($y_0 = f(x_0)$ и $y_1 = f(x_1)$). Этим точкам соответствуют значения 0 и 1 переменной q . Следовательно,

$$\int_{x_0}^{x_1} f(x) dx \approx \int_0^1 (y_0 + q\Delta y_0) h dq = h \left[y_0 q + \frac{q^2}{2} \Delta y_0 \right]_0^1 = \\ = h \left(y_0 + \frac{y_1 - y_0}{2} \right) = h \frac{y_0 + y_1}{2}. \quad (5.17)$$

Получена **простейшая квадратурная формула трапеций**, к которой легко прийти и из геометрических соображений (рис. 5.3).

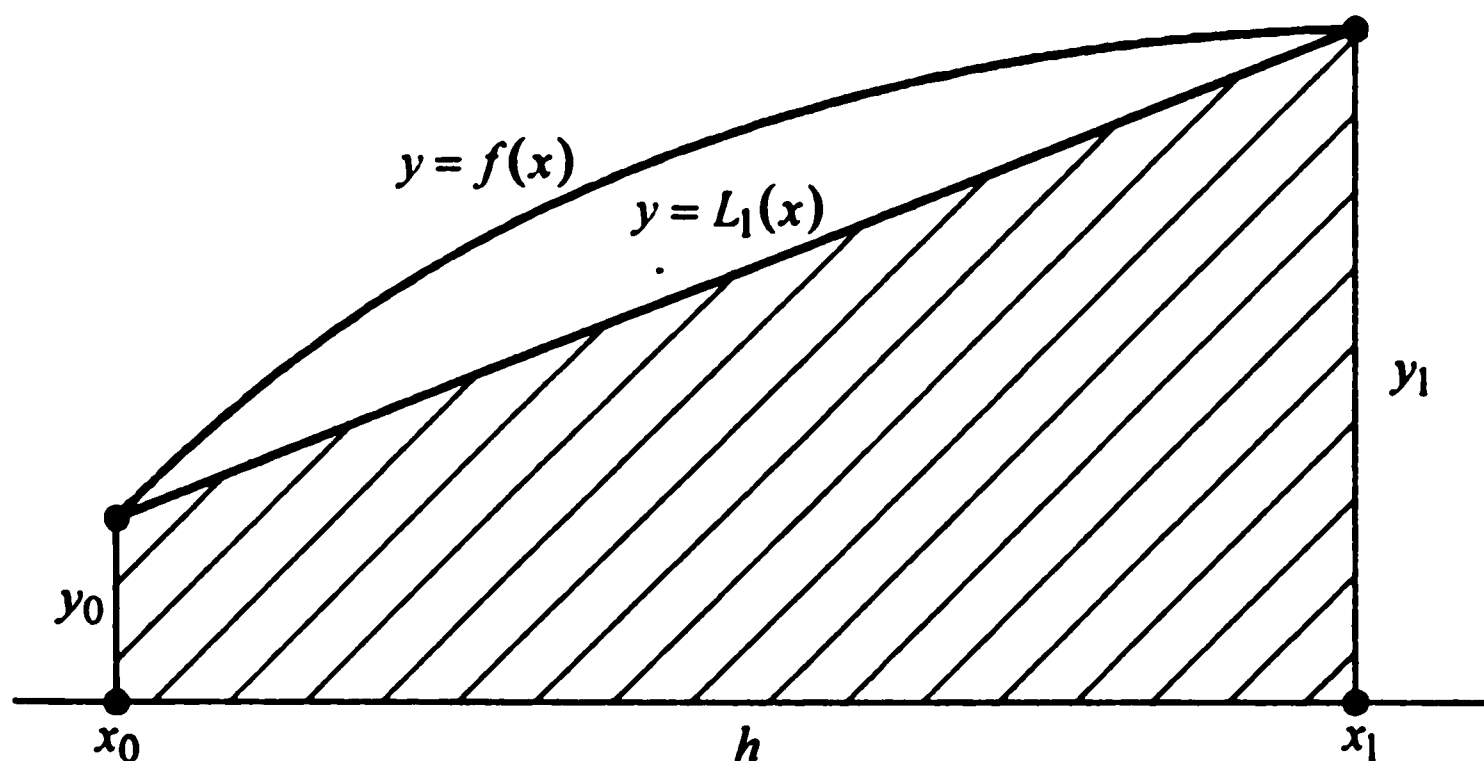


Рис. 5.3. Геометрическая интерпретация простейшей формулы трапеций

Остаточный член этой формулы найдем интегрированием остаточного члена $R_1(x)$ формулы линейной интерполяции, преобразованного к виду

$$R_1(x_0 + qh) = \frac{f''(\xi)}{2} h^2 q(q-1).$$

Имеем:

$$\begin{aligned} r_1 &:= \int_{x_0}^{x_1} f(x) dx - \frac{h}{2}(y_0 + y_1) = \frac{h^3}{2} \int_0^1 f''(\xi)(q^2 - q) dq = \\ &= \frac{h^3}{2} f''(\xi_1) \int_0^1 (q^2 - q) dq = -\frac{f''(\xi_1)}{12} h^3. \end{aligned} \quad (5.18)$$

В найденном выражении остаточного члена $\xi_1 \in (x_0, x_1)$ — некоторая точка, отвечающая упоминавшейся в предыдущем параграфе интегральной теореме о среднем, которая здесь также применима, в силу неположительности функции $q^2 - q$ на отрезке $[0, 1]$.

2) Положим в (5.16) $n = 2$, т.е. проинтерполируем функцию $f(x)$ по трем точкам: x_0 , $x_1 = x_0 + h$ и $x_2 = x_0 + 2h$. Тогда

$$\begin{aligned} \int_{x_0}^{x_2} f(x) dx &\approx \int_0^2 \left[y_0 + q\Delta y_0 + \frac{q(q-1)}{2} \Delta^2 y_0 \right] h dq = \\ &= h \left[2y_0 + 2(y_1 - y_0) + \frac{1}{3}(y_2 - 2y_1 + y_0) \right] = \frac{h}{3}(y_0 + 4y_1 + y_2). \end{aligned} \quad (5.19)$$

Полученное приближенное равенство назовем *простейшей формулой Симпсона*^{*)}.

Поставим задачу найти *остаточный член*

$$r_2 := \int_{x_0}^{x_2} f(x) dx - \frac{h}{3}(y_0 + 4y_1 + y_2) \quad (5.20)$$

простейшей формулы Симпсона (5.19). Так как функция $q(q-1)(q-2)$ меняет знак на промежутке $[0, 2]$, то здесь нельзя воспользоваться интегральной теоремой о среднем при интегрировании остаточного члена

$$R_2(x_0 + qh) = \frac{f'''(\xi)}{3!} h^3 q(q-1)(q-2)$$

формулы квадратичной интерполяции, как это было в случае 1). Поэтому для выявления вида r_2 выберем искусственный путь.

Для удобства рассмотрим применение простейшей квадратурной формулы Симпсона (5.19) к интегралу с симметричными границами:

$$\int_{-h}^h f(x) dx = \frac{h}{3}[f(-h) + 4f(0) + f(h)] + r(h).$$

^{*)} Симпсон Томас (1710–1761) — английский математик; прошел путь от ткача до профессора военной академии.

При любом $t \in [0, h]$ ее остаточный член есть

$$r(t) := \int_{-t}^t f(\tau) d\tau - \frac{t}{3} [f(-t) + 4f(0) + f(t)]. \quad (5.21)$$

Введем в рассмотрение функцию

$$v(t) := r(t) - \left(\frac{t}{h}\right)^5 r(h) \quad (5.22)$$

и изучим поведение $v(t)$ и нескольких ее первых производных, предполагая, что исходная функция $f(x)$ четырежды дифференцируема на $[-h, h]$.

Так как $\frac{d}{dt} \int_{-t}^t f(\tau) d\tau = f(t) - (-1)f(-t)$, то

$$v'(t) = f(t) + f(-t) - \frac{1}{3} [f(-t) + 4f(0) + f(t)] - \\ - \frac{t}{3} [f'(t) - f'(-t)] - \frac{5t^4}{h^5} r(h).$$

Дальнейшее дифференцирование дает:

$$v''(t) = \frac{1}{3} f'(t) - \frac{1}{3} f'(-t) - \frac{t}{3} [f''(t) + f''(-t)] - \frac{20t^3}{h^5} r(h), \\ v'''(t) = -\frac{t}{3} [f'''(t) - f'''(-t)] - \frac{60t^2}{h^5} r(h).$$

Последнее, благодаря формуле конечных приращений Лагранжа, примененной к разности третьих производных (в квадратных скобках), можно переписать в виде

$$v'''(t) = -\frac{2}{3} t^2 \left[f'''(\xi) + \frac{90}{h^5} r(h) \right], \quad (5.23)$$

где ξ — некоторая точка из интервала $(-t, t)$.

Теперь обратимся к анализу функции $v(t)$ и ее производных.

Как видно из (5.21), $r(0) = 0$; поэтому и $v(0) = 0$. Подстановка $t = h$ в (5.22) также приводит к нулевому значению $v(t)$. Следовательно, к функции $v(t)$ на отрезке $[0, h]$ применима теорема Ролля, согласно которой существует такая точка $t_1 \in (0, h)$, что $v'(t_1) = 0$ (рис. 5.4). Непосредственной подстановкой значения $t = 0$ в выражение $v'(t)$ убеждаемся, что $v'(0) = 0$. Значит, теорема Ролля применима и к функции $v'(t)$ на отрезке $[0, t_1]$,

т.е.

$$\exists t_2 \in (0, t_1): v''(t_2) = 0.$$

Так как $v''(0) = 0$, то по той же теореме

$$\exists \Theta \in (0, t_2): v'''(\Theta) = 0.$$

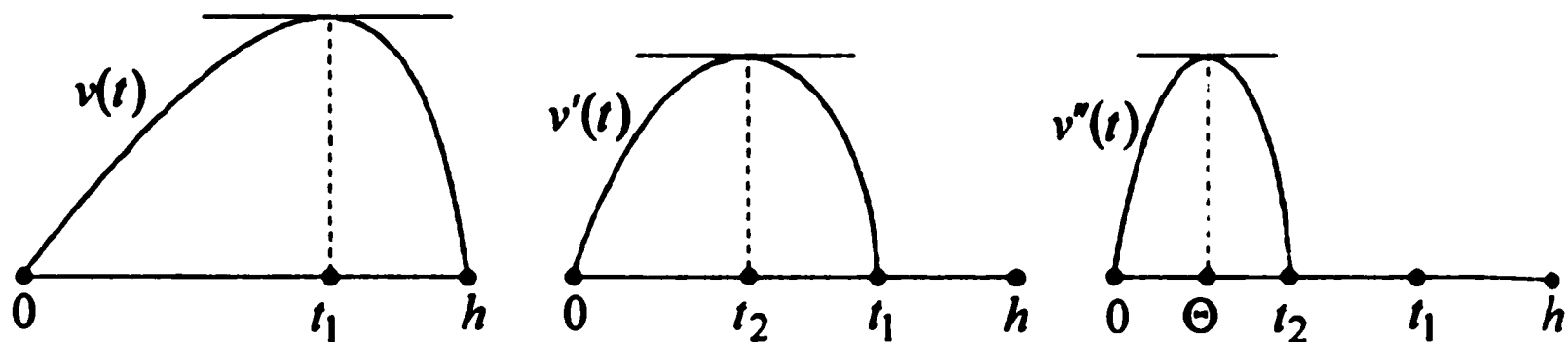


Рис. 5.4. Геометрическая интерпретация трехкратного применения теоремы Ролля при выводе остаточного члена простейшей формулы Симпсона

Таким образом, в соответствии с (5.23), при некотором $\Theta \in (0, t_2) \subset (0, h)$ справедливо равенство

$$-\frac{2}{3}\Theta^2 \left[f^{IV}(\xi) + \frac{90}{h^5}r(h) \right] = 0,$$

из которого получаем выражение остаточного члена

$$r(h) = -\frac{h^5}{90} f^{IV}(\xi), \quad \xi \in (-h, h).$$

Очевидно, эта формула может быть отнесена и к выражению r_2 , определенному в (5.20), где промежуток интегрирования $[x_0, x_2]$ следует рассматривать как симметричный относительно точки x_1 : $x \in [x_1 - h, x_1 + h]$; т.е. можно записать

$$r_2 = -\frac{h^5}{90} f^{IV}(\xi), \quad \xi \in (x_0, x_2). \quad (5.24)$$

3) Фиксируя степень $n = k$ интерполяционного многочлена (5.16) равной 3, 4, 5 и т.д., придем к частным формулам Ньютона–Котеса, подобным полученным выше простейшим формулам трапеций и Симпсона. Представим все эти частные случаи, включая уже рассмотренные, формулой вида

$$\int_{x_0}^{x_k} f(x) dx = B_k h \sum_{i=0}^k a_i^{(k)} f(x_i) + r_k(h), \quad (5.25)$$

где по-прежнему считается $x_i = x_0 + ih$, а коэффициенты B_k , $a_i^{(k)}$ и остаточные члены $r_k(h)$ задаются таблицей 5.1 (точка $\xi \in (x_0, x_k)$, разумеется, для каждого k своя).

Т а б л и ц а 5.1

Параметры некоторых частных формул
Ньютона–Котеса вида (5.25) [90]

k	B_k	$a_0^{(k)}$	$a_1^{(k)}$	$a_2^{(k)}$	$a_3^{(k)}$	$a_4^{(k)}$	$a_5^{(k)}$	...	$r_k(h)$
1	$\frac{1}{2}$	1	1						$-\frac{h^3}{12}f''(\xi)$
2	$\frac{1}{3}$	1	4	1					$-\frac{h^5}{90}f^{IV}(\xi)$
3	$\frac{3}{8}$	1	3	3	1				$-\frac{3h^5}{80}f^{IV}(\xi)$
4	$\frac{2}{45}$	7	32	12	32	7			$-\frac{8}{945}h^7f^{VI}(\xi)$
5	$\frac{5}{288}$	19	75	50	50	75	19		$-\frac{275}{12096}h^7f^{VI}(\xi)$
...	...	...	...	...	...	...	...	...	...

Замечание 5.1. В отличие от рассмотренных представлений формулы Ньютона–Котеса (5.14) или (5.25) непосредственно через значения подынтегральной функции, имеется другой ее вид, который можно назвать конечноразностным и который известен как *формула Грегори*^{*)} [11, 90]:

$$\int_{x_0}^{x_0+kh} f(x) dx \approx h \left[\left(\frac{1}{2} y_0 + y_1 + \dots + y_{k-1} + \frac{1}{2} y_k \right) + \frac{1}{12} (\Delta y_0 - \Delta y_{k-1}) - \right. \\ \left. - \frac{1}{24} (\Delta^2 y_0 - \Delta^2 y_{k-2}) + \frac{19}{720} (\Delta^3 y_0 - \Delta^3 y_{k-3}) - \frac{3}{160} (\Delta^4 y_0 - \Delta^4 y_{k-4}) + \dots \right].$$

Очевидно, эта формула имеет такие же достоинства, как и конечноразностные интерполяционные формулы (см. § 1.5): слагаемые здесь расположены в порядке убывания значимости. При этом фиксирование $k = 1, 2, \dots$ определяет соответственно формулы трапеций, Симпсона и др.

^{*)} Грегори Джеймс (1638–1675) — шотландский математик и астроном.

§ 5.3. СОСТАВНЫЕ КВАДРАТУРНЫЕ ФОРМУЛЫ ТРАПЕЦИЙ И СИМПСОНА

Применение формул Ньютона–Котеса высоких порядков, т.е. при относительно больших значениях параметра $k \in \mathbb{N}$ в (5.25), может быть оправданным лишь при достаточно высокой гладкости подынтегральной функции $f(x)$. Более употребительными являются квадратурные правила, получающиеся путем дробления промежутка интегрирования на большое число мелких частей, интегрирование на каждой из которых производится с помощью однотипных простейших формул невысокого порядка. Выведем два таких правила — трапеций и Симпсона.

1) Прежде всего, заметим, что простейшая формула трапеций (5.17) с остаточным членом (5.18) применительно к интегрированию по отрезку $[x_{i-1}, x_i]$ может быть записана в виде точного равенства

$$\int_{x_{i-1}}^{x_i} f(x) dx = \frac{y_{i-1} + y_i}{2} h - \frac{f''(\xi_i)}{12} h^3, \quad (5.26)$$

где ξ_i — некоторая, вообще говоря, неизвестная точка интервала (x_{i-1}, x_i) , а y_i , как и прежде, есть укороченная запись значения $f(x_i)$.

Выполняя разбиение (5.4) исходного промежутка интегрирования $[a, b]$ на n частей с шагом $h = \frac{b-a}{n}$ и применяя к каждой из частей, на которые по свойству аддитивности расчленяется исходный интеграл, формулу (5.26), будем иметь

$$\int_a^b f(x) dx = \sum_{i=1}^n \int_{x_{i-1}}^{x_i} f(x) dx = \frac{h}{2} \sum_{i=1}^n (y_{i-1} + y_i) - \frac{h^3}{12} \sum_{i=1}^n f''(\xi_i). \quad (5.27)$$

Отсюда следует, что искомое значение интеграла можно приближенно найти по формуле

$$I \approx I^T := h \left(\frac{y_0 + y_n}{2} + y_1 + y_2 + \dots + y_{n-1} \right), \quad (5.28)$$

которую в дальнейшем будем называть просто **формулой трапеций**, а погрешность приближенного равенства (5.28) можно охарактеризовать остаточным членом r^T , полученным упрощением последнего слагаемого в (5.27). По обобщенной теореме о среднем значении функции на отрезке существует такая точка $\xi_T \in (a, b)$, что

$$\sum_{i=1}^n h f''(\xi_i) = f''(\xi_T)(b-a)$$

(площадь ступенчатой фигуры, составленной из прямоугольников с основаниями h и высотами $f''(\xi_i)$, можно отождествить с площадью одного прямоугольника с основанием $\sum_{i=1}^n h = b - a$ и высотой $f''(\xi_T)$). Таким образом, *остаточный член формулы трапеций* (5.28) есть

$$r^T := I - I^T = -\frac{b-a}{12} h^2 f''(\xi_i), \quad \xi_i \in (a, b). \quad (5.29)$$

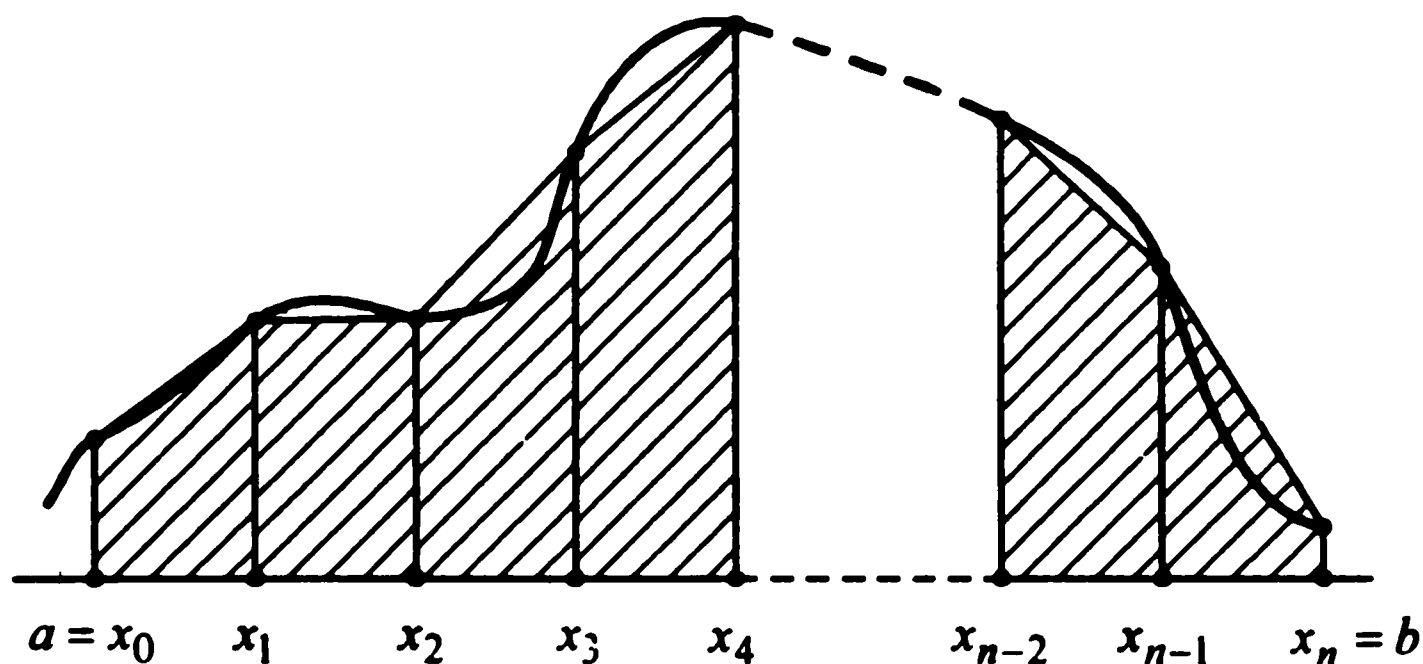


Рис. 5.5. Геометрическая интерпретация составной квадратурной формулы трапеций (5.28)

2) Аналогично равенству (5.26) на основе простейшей формулы Симпсона (5.19) и ее остаточного члена (5.24) запишем равенство

$$\int_{x_{2i-2}}^{x_{2i}} f(x) dx = \frac{h}{3} (y_{2i-2} + 4y_{2i-1} + y_{2i}) - \frac{h^5}{90} f^{(4)}(\xi_i), \quad (5.30)$$

$$\xi_i \in (x_{2i-2}, x_{2i}).$$

Выполнив разбиение (5.4) так, чтобы число элементарных промежутков $n = 2m$ было четным, исходный интеграл представим суммой m интегралов вида (5.30):

$$\int_a^b f(x) dx = \sum_{i=1}^m \int_{x_{2i-2}}^{x_{2i}} f(x) dx = \frac{h}{3} \sum_{i=1}^m (y_{2i-2} + 4y_{2i-1} + y_{2i}) - \frac{h^5}{90} \sum_{i=1}^m f^{(4)}(\xi_i).$$

Отсюда получается формула численного интегрирования

$$I \approx I^C := \frac{h}{3}(y_0 + y_{2m} + 4\sigma_1 + 2\sigma_2), \quad (5.31)$$

где $\sigma_1 := y_1 + y_3 + \dots + y_{2m-1}$, $\sigma_2 := y_2 + y_4 + \dots + y_{2m-2}$, которая впредь будет называться **формулой Симпсона**, и **остаточный член** ее

$$r^C := I - I^C = -\frac{h^4}{180} \sum_{i=1}^m 2h f^{(4)}(\xi_i) = -\frac{b-a}{180} h^4 f^{(4)}(\xi_C) \quad (5.32)$$

с некоторой точкой ξ_C из интервала (a, b) .

Подводя итоги этого параграфа, можно сказать, что замена подынтегральной функции $f(x)$ на промежутке интегрирования $[a, b]$, разбитом на n частей с шагом h , кусочно-линейной функцией (рис. 5.5) приводит к его приближенному значению I' (5.28) с ошибкой, убывающей при $h \rightarrow 0$ (согласно (5.29)) по квадратичному закону. Если же сделать кусочно-квадратичную интерполяцию подынтегральной функции по сдвоенным элементарным промежуткам $[x_0, x_2]$, $[x_2, x_4]$, ..., $[x_{2m-2}, x_{2m}]$, то ошибка получаемого таким образом приближенного равенства $I \approx I^C$ (5.31) в соответствии с (5.32) будет убывать пропорционально уже четвертой степени h . Это обстоятельство обусловило чрезвычайную популярность формулы Симпсона, ибо повышение порядка точности интерполяции всего на одну единицу повлекло повышение точности интегрирования на два порядка (при сохранении простоты расчетной формулы). Посмотрев на строку табл.5.1, соответствующую значению $k=3$, легко понять, что применение кусочно-кубической интерполяции на основе формулы Ньютона–Котеса не даст принципиального улучшения качества соответствующей формулы численного интегрирования по сравнению с формулой Симпсона.

Пример 5.1. Оценим, какую точность можно гарантировать при вычислении интеграла

$$I = \int_0^1 e^{-x^2} dx$$

по формулам трапеций и Симпсона при разбиении промежутка интегрирования на восемь частей.

Выполнив последовательное дифференцирование функции $f(x) = e^{-x^2}$, находим

$$M_2 := \max_{x \in [a, b]} |f''(x)| = \max_{x \in [0, 1]} |2e^{-x^2}(2x^2 - 1)| = 2$$

и

$$M_4 := \max_{x \in [a, b]} |f^{(4)}(x)| = \max_{x \in [0, 1]} |4e^{-x^2}(4x^4 - 12x^2 + 3)| = 12.$$

Теперь, учитывая, что $b - a = 1$, $h = \frac{b-a}{n} = \frac{1}{8}$, в соответствии с формулами (5.29) и (5.32) получаем требуемые оценки:

$$|I - I^T| \leq \frac{M_2(b-a)}{12} h^2 = \frac{2}{12} \left(\frac{1}{8}\right)^2 = \frac{1}{384} < 0.003$$

и

$$|I - I^C| \leq \frac{M_4(b-a)}{180} h^4 = \frac{12}{180} \left(\frac{1}{8}\right)^4 = \frac{1}{61440} < 0.00002.$$

§ 5.4. СООТНОШЕНИЯ МЕЖДУ ФОРМУЛАМИ ПРЯМОУГОЛЬНИКОВ, ТРАПЕЦИЙ И СИМПСОНА

Воспользоваться оценками квадратурных формул подобно тому, как это было сделано в примере 5.1, удастся крайне редко. Поэтому вычисление интегралов с нужной точностью обычно производят посредством последовательного дробления шага (как правило, делением пополам) до выполнения некоторых критериев точности. Обоснование таких критериев и конкретные алгоритмы будут приведены в следующем параграфе. Здесь же изучим связи, которые возникают между приближенными значениями I^P , I^T и I^C интеграла I в процессе его вычисления по формулам прямоугольников, трапеций и Симпсона с шагом h и шагом $H = 2h$ (рис. 5.6).

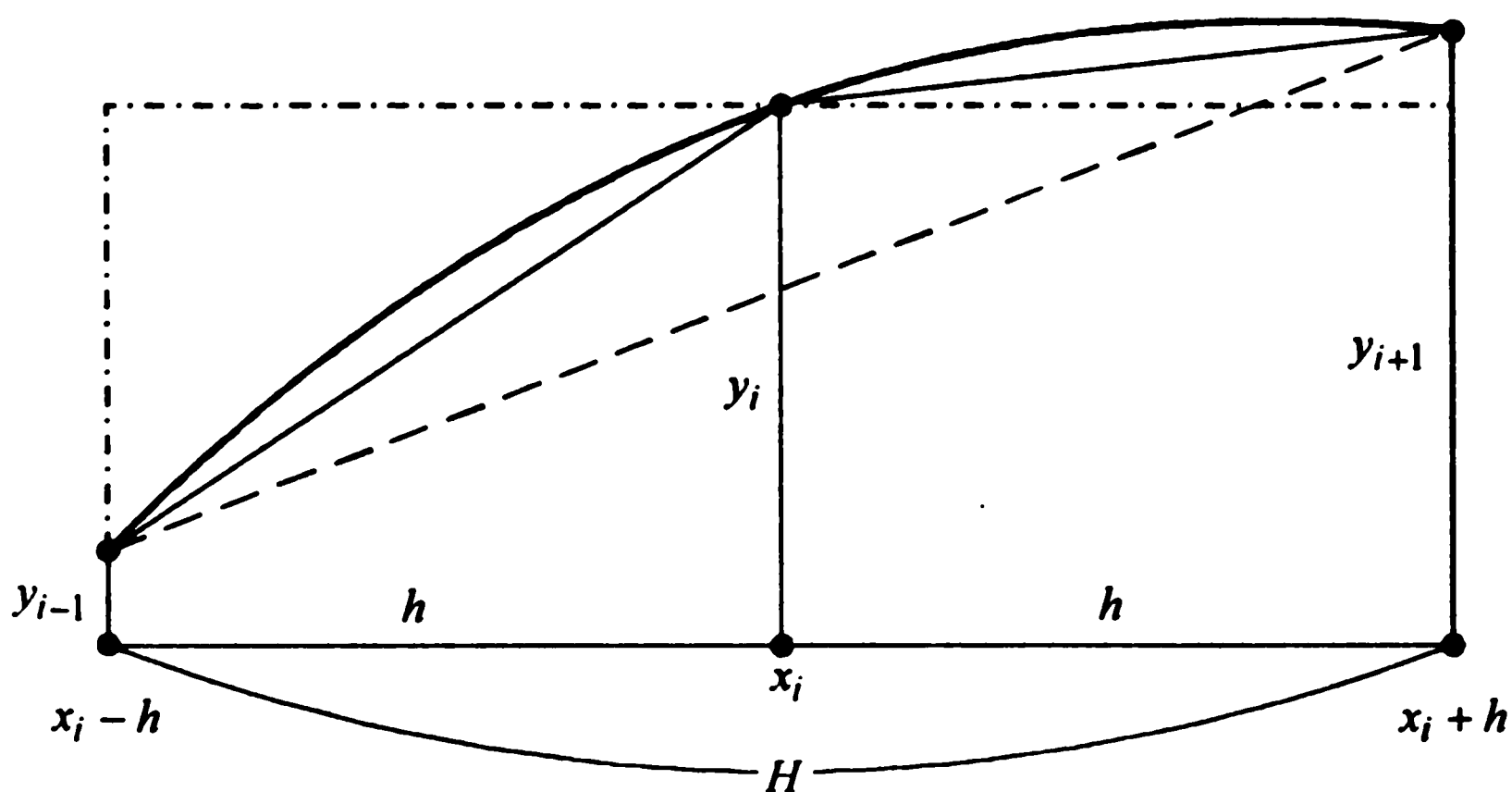


Рис. 5.6. К выводу соотношений между $I_i^P(H)$, $I_i^T(H)$, $I_i^T(h)$ и $I_i^C(h)$

Рассмотрим сначала приближения к интегралу

$$I_i := \int_{x_i-h}^{x_i+h} f(x) dx,$$

полученные по формуле прямоугольников с шагом H ($I_i^n(H)$), по формуле трапеций с шагами H и $h = \frac{H}{2}$ ($I_i^T(H)$ и $I_i^T(h)$), а также по формуле Симпсона с шагом h ($I_i^C(h)$), через значения $y_{i-1} = f(x_i - h)$, $y_i = f(x_i)$ и $y_{i+1} = f(x_i + h)$.

Имеем:

$$I_i^n(H) = Hy_i = 2hy_i;$$

$$I_i^T(H) = \frac{y_{i-1} + y_{i+1}}{2} H = h(y_{i-1} + y_{i+1});$$

$$I_i^T(h) = \frac{y_{i-1} + y_i}{2} h + \frac{y_i + y_{i+1}}{2} h = \frac{h}{2}(y_{i-1} + 2y_i + y_{i+1});$$

$$I_i^C(h) = \frac{h}{3}(y_{i-1} + 4y_i + y_{i+1}).$$

Беря полусумму правых частей первых двух из этих равенств, получаем правую часть третьего. Следовательно,

$$I_i^T(h) = \frac{1}{2}(I_i^n(H) + I_i^T(H)). \quad (5.33)$$

Если же взять среднее взвешенное этих же равенств с весами $\frac{2}{3}$ и $\frac{1}{3}$ соответственно, то придем к четвертому равенству, т.е.

$$I_i^C(h) = \frac{2}{3}I_i^n(H) + \frac{1}{3}I_i^T(H). \quad (5.34)$$

Поскольку исходный интеграл есть $I = \sum_{i=1}^m I_i$, где $m = \frac{b-a}{H}$, то индекс i в соотношениях (5.33), (5.34) можно отбросить.

Итак, если $h = \frac{H}{2}$, то

$$I^T(h) = \frac{1}{2}(I^n(H) + I^T(H)), \quad (5.35)$$

$$I^C(h) = \frac{2}{3}I^n(H) + \frac{1}{3}I^T(H). \quad (5.36)$$

На базе этих связей можно строить эффективные алгоритмы вычисления интегралов. Суть в том, что, производя для уточнения значения интеграла дробление шага H пополам, мы должны вычислять новые значения подынтегральной функции в точках, расположенных посередине предыдущих элементарных отрезков. Эти новые значения используются для вычисления значения $I^P(H)$, все остальные значения функции передаются на этот этап с предыдущего уже в просуммированном виде, т.е. как $I^T(H)$. Так можно получать либо более точное, чем $I^T(H)$, значение $I^T(h)$, либо $I^C(h)$.

Заметим, кстати, что значение $I^C(h)$ можно вычислить также по формуле

$$I^C(h) = I^T(h) + R^T(h), \quad (5.37)$$

где $R^T(h) := \frac{1}{3}(I^T(h) - I^T(H))$ — так называемая *поправка Ричардсона*^{*)}. Действительно, возвращаясь к интегралу I_i , непосредственно (справа налево) проверяем справедливость i -го фрагмента формулы (5.37):

$$\begin{aligned} I_i^T(h) + R_i^T(h) &= I_i^T(h) + \frac{1}{3}(I_i^T(h) - I_i^T(H)) = \frac{4}{3}I_i^T(h) - \frac{1}{3}I_i^T(H) = \\ &= \frac{4}{3} \cdot \frac{h}{2}(y_{i-1} + 2y_i + y_{i+1}) - \frac{1}{3} \cdot h(y_{i-1} + y_{i+1}) = \frac{h}{3}(y_{i-1} + 4y_i + y_{i+1}) = \\ &= I_i^C(h); \end{aligned}$$

из аддитивности интеграла теперь следует (5.37).

Таким образом, вычисление поправки Ричардсона позволяет уточнить приближенное значение $I^T(h)$ интеграла I , полученное по формуле трапеций. Другое ее назначение будет понятно из материала следующего параграфа.

§ 5.5. ПРИНЦИП РУНГЕ ПРАКТИЧЕСКОГО ОЦЕНИВАНИЯ ПОГРЕШНОСТЕЙ. АЛГОРИТМ РОМБЕРГА

Пусть к приближенному вычислению значения I данного интеграла применяется некая квадратурная формула p -го порядка точности I^P из семейства составных формул Ньютона–Котеса. При условии непрерывности p -ой производной

^{*)} Ричардсон Льюис Фрай (1881-1953) — английский геофизик.

подынтегральной функции это означает существование такой постоянной C , что

$$I = I^P(h) + Ch^P. \quad (5.38)$$

При уменьшении вдвое шага h численного интегрирования по той же формуле p -го порядка можно записать такое же равенство, но с другой постоянной C_1 :

$$I = I^P\left(\frac{h}{2}\right) + C_1\left(\frac{h}{2}\right)^P. \quad (5.39)$$

Считая, что при малом h постоянные C и C_1 близки, из (5.38) и (5.39) имеем

$$I^P(h) + Ch^P = I^P\left(\frac{h}{2}\right) + C_1\left(\frac{h}{2}\right)^P \approx I^P\left(\frac{h}{2}\right) + C\left(\frac{h}{2}\right)^P$$

и, следовательно,

$$C \approx C_1 \approx \frac{I^P\left(\frac{h}{2}\right) - I^P(h)}{h^P - \left(\frac{h}{2}\right)^P}.$$

Подставив это значение C_1 в (5.39), приходим к выражению

$$I \approx I^P\left(\frac{h}{2}\right) + \frac{I^P\left(\frac{h}{2}\right) - I^P(h)}{2^P - 1}. \quad (5.40)$$

К приближенному равенству (5.40) можно относиться двояко. С одной стороны, переписав его в виде

$$I - I^P\left(\frac{h}{2}\right) \approx \frac{I^P\left(\frac{h}{2}\right) - I^P(h)}{2^P - 1}, \quad (5.41)$$

получаем возможность хотя бы грубо контролировать точность численного интегрирования на основе двойного счета (с шагом h

и с шагом $\frac{h}{2}$). В этом и заключается широко применяемый **принцип Рунге** практического оценивания погрешностей. Его применение считается правомочным [21], если выполняется неравенство

$$\left| 2^P \cdot \frac{I^P\left(\frac{h}{2}\right) - I^P(h)}{I^P(h) - I^P(2h)} - 1 \right| < 0.1.$$

С другой стороны, второе слагаемое в формуле (5.40) позволяет уточнить «дешевым» способом приближенное значение $I^p\left(\frac{h}{2}\right)$ интеграла I . Заметим, что при $p = 2$ формула (5.38) есть формула трапеций

$$I = I^T(h) + Ch^2,$$

и выражение (5.41) в этом случае совпадает с определенной в предыдущем параграфе поправкой Ричардсона $R^T\left(\frac{h}{2}\right)$, а правая

часть формулы (5.40) есть значение $I^C\left(\frac{h}{2}\right)$, соответствующее формуле Симпсона. Естественно назвать **обобщенной поправкой Ричардсона** получаемую с помощью двойного счета величину^{*)}

$$R^p\left(\frac{h}{2}\right) := \frac{I^p\left(\frac{h}{2}\right) - I^p(h)}{2^p - 1}.$$

Итогом рассуждений предыдущего и настоящего параграфов может служить, например, следующий **алгоритм прямоугольников-трапеций** вычисления интеграла I с заданной точностью ε .

Шаг 1. Полагаем $n := 1$, $H := b - a$, $I^T(H) := \frac{H}{2}[f(a) + f(b)]$.

Шаг 2. Вычисляем $h = \frac{H}{2}$, $x_1 = a + h$, $x_i = x_{i-1} + h$

$$(i = 2, \dots, n), \quad y_i = f(x_i) \quad (i = 1, 2, \dots, n), \quad I^n(H) = H \sum_{i=1}^n y_i.$$

Шаг 3. Вычисляем $I^T(h) = \frac{1}{2}[I^n(H) + I^T(H)]$,

$$R^T(h) = \frac{1}{3}[I^T(h) - I^T(H)].$$

Шаг 4. Сравниваем $|R^T(h)|$ с ε . Если $|R^T(h)| > \varepsilon$, то полагаем $n := 2n$, $H := h$, $I^T(H) := I^T(h)$ и возвращаемся к шагу 2.

Шаг 5. Вычисляем $I^C(h) = I^T(h) + R^T(h)$ и принимаем $I \approx I^C(h)$.

^{*)} Не вызовет затруднений рассмотрение и более общего случая, когда шаг h заменяется на шаг $h_1 = \gamma h$, где $\gamma \in (0, 1)$.

Если есть основания считать, что подынтегральная функция обладает достаточно высокой степенью гладкости, то для вычисления интеграла можно применить алгоритм, предложенный шведским математиком Ромбергом. Базируется этот алгоритм на связях между составными квадратурными формулами Ньютона–Котеса четных порядков типа установленной выше связи (5.37) между формулами трапеций и Симпсона через поправку Ричардсона. Так, можно показать, что правая часть соответствующего (5.40) при $p = 4$ приближенного равенства

$$I \approx I^4\left(\frac{h}{2}\right) + \frac{I^4\left(\frac{h}{2}\right) - I^4(h)}{15}$$

(где $I^4(h) \equiv I^C(h)$, согласно предыдущему) совпадает со значением $I^6\left(\frac{h}{2}\right)$, которое могло бы быть подсчитано по составной формуле Ньютона–Котеса, отвечающей строке $k = 4$ в табл.5.1, и т.д.

Следовательно, заложив процесс деления шага пополам и выполнив первоначальный подсчет приближенного значения интеграла I по формуле трапеций, дальнейшие его уточнения можно производить рекуррентно путем прибавления обобщенных поправок Ричардсона к предыдущим приближениям к I .

Таким образом, алгоритм Ромберга определяется следующей совокупностью формул [9]:

$$h_0 := b - a, \quad h_i = \frac{h_{i-1}}{2}, \quad I^{(0)}(h_0) := I^T(h_0), \quad I^{(0)}(h_i) := I^T(h_i),$$

$$R^{(k-1)}(h_i) := \frac{1}{2^{2k} - 1} [I^{(k-1)}(h_i) - I^{(k-1)}(h_{i-1})],$$

$$I^{(k)}(h_i) = I^{(k-1)}(h_i) + R^{(k-1)}(h_i),$$

где i принимает значения $1, 2, 3, \dots$, а k изменяется от 1 до i ; счет заканчивается при выполнении условия $|R^{(k-1)}(h_i)| \leq \varepsilon$, после чего полагают $I \approx I^{(k)}(h_i)$.

Последовательность уточнений значения интеграла в алгоритме Ромберга удобно проследить по таблице 5.2 (см. нумерацию клеток таблицы).

Схема последовательных уточнений интеграла алгоритмом Ромберга

i	h_i	$k = 0$	$k = 1$	$k = 2$	$k = 3$	$k = 4$	...
0	h_0	1 $I^{(0)}(h_0)$					
1	h_1	2 $I^{(0)}(h_1)$	3 $I^{(1)}(h_1)$				
2	h_2	4 $I^{(0)}(h_2)$	5 $I^{(1)}(h_2)$	6 $I^{(2)}(h_2)$			
3	h_3	7 $I^{(0)}(h_3)$	8 $I^{(1)}(h_3)$	9 $I^{(2)}(h_3)$	10 $I^{(3)}(h_3)$		
4	h_4	11 $I^{(0)}(h_4)$	12 $I^{(1)}(h_4)$	13 $I^{(2)}(h_4)$	14 $I^{(3)}(h_4)$	15 $I^{(4)}(h_4)$	
...	...	...	...	...	...	...	...

Наверное, не будет лишним обратить внимание на формирование столбца этой таблицы, соответствующего $k = 0$. Здесь должны находиться значения интеграла, получаемые по формуле трапеций с уменьшающимся в два раза шагом. Очевидно, к этому этапу целесообразно подключить часть приведенного выше алгоритма прямоугольников-трапеций, которая позволит экономично вычислять значения $I^{(0)}(h_i) = I^T(h_i)$ на основе соотношения (5.35).

§ 5.6. КВАДРАТУРНЫЕ ФОРМУЛЫ ЧЕБЫШЕВА И ГАУССА

Общий вид линейной квадратурной формулы — это

$$\int\limits_a^b f(x) dx \approx \sum\limits_i A_i f(x_i),$$

(5.42)

где фиксированные аргументы x_i называют узлами, а коэффициенты A_i — весами (весовыми коэффициентами) квадратурной формулы (определенный интеграл приближенно равен среднему взвешенному значений подынтегральной функции, вычисленных в определенных точках промежутка интегрирования).

Все рассмотренные выше квадратурные формулы характерны тем, что узлы в них брались равноотстоящими с шагом h , а веса находились в результате подмены подынтегральной функции $f(x)$ кусочно-постоянной в случае формул прямоугольников,

кусочно-линейной в случае формулы трапеций, кусочно-квадратичной в случае формулы Симпсона и т.д. Например, у составной формулы трапеций набор весов получился следующий:

$$\frac{h}{2}, h, h, \dots, h, \frac{h}{2},$$

а у составной формулы Симпсона —

$$\frac{h}{3}, \frac{4h}{3}, \frac{2h}{3}, \frac{4h}{3}, \frac{2h}{3}, \dots, \frac{4h}{3}, \frac{h}{3}.$$

Далее мы откажемся от равномерного распределения узлов x_i на промежутке интегрирования $[a, b]$ и будем их находить из тех или иных соображений. В таком случае целесообразно предварительно сделать линейную замену (см. § 2.2)

$$x = \frac{a+b}{2} + \frac{b-a}{2}t$$

и преобразовать исходный интеграл к интегралу со стандартным промежутком интегрирования $[-1, 1]$:

$$\int_a^b f(x)dx = \frac{b-a}{2} \int_{-1}^1 f\left(\frac{a+b}{2} + \frac{b-a}{2}t\right)dt. \quad (5.43)$$

Равенство (5.43) позволяет без ограничения общности рассматривать теперь вычисление интеграла

$$\hat{I} \approx \int_{-1}^1 \varphi(t)dt,$$

т.е. строить квадратурные формулы вида

$$\hat{I} \approx \sum_{i=1}^n A_i \varphi(t_i), \quad (5.44)$$

от которых на основе (5.43) легко перейти к квадратурным формулам (5.42).

Сделаем еще два предварительных замечания.

Во-первых, обратим внимание на разную роль параметра n в квадратурных формулах предыдущих параграфов и в формуле (5.44): если ранее n означало число элементарных отрезков, на которые разбивался промежуток интегрирования, то здесь n — это число узлов.

Во-вторых, квадратурные формулы, основанные на равномерном разбиении отрезка $[a, b]$ на n частей с шагом h , сравнивали по точности в зависимости от степени h , содержащейся в выражениях их остаточных членов (см. (5.12), (5.29), (5.32) и табл.5.1). Но те же остаточные члены позволяют судить о точности иначе. Поскольку в выражение остаточного члена квадратурной

формулы входит множителем производная подынтегральной функции определенного, например, k -го порядка, и поскольку k -я производная многочлена $(k-1)$ -й степени равна нулю, можно сказать, что соответствующая квадратурная формула точна для многочленов степени $k-1$. В таком случае $k-1$ есть **алгебраический порядок точности** квадратурной формулы.

Потребуем, чтобы в формуле (5.44) значения функции $\varphi(t_i)$ суммировались с одинаковыми весами $A_i \equiv A$, а узлы t_i получающейся при этом формулы

$$\int_{-1}^1 \varphi(t) dt \approx A \sum_{i=1}^n \varphi(t_i) \quad (5.45)$$

располагались на отрезке $[-1, 1]$ так, чтобы формула (5.45) была точна для многочленов степени n . Отвечающая этим условиям формула называется **квадратурной формулой Чебышева** (она хороша тем, что минимизирует вероятную ошибку, если значения $\varphi(t_i)$ имеют нормально распределенную случайную ошибку [46]).

Считая, что равенство (5.45) должно быть непременно точным для постоянной функции, подставим в него $\varphi(t) \equiv 1$. Имеем

$$\int_{-1}^1 1 dt = A \sum_{i=1}^n 1, \quad \text{откуда} \quad A = \frac{2}{n}.$$

Следовательно, вместо (5.45) теперь можно использовать более конкретный шаблон квадратурной формулы Чебышева

$$\int_{-1}^1 \varphi(t) dt \approx \frac{2}{n} \sum_{i=1}^n \varphi(t_i). \quad (5.46)$$

Чтобы выписать совокупность условий на расположение узлов t_i на отрезке $[-1, 1]$, можно либо подставить в левую и правую части равенства (5.46) функцию-многочлен

$$\varphi(t) = a_0 + a_1 t + a_2 t^2 + \dots + a_n t^n$$

и считать это равенство точным при любых a_i , либо поочередно подставлять в (5.46) представителей многочленов, коими являются степенные функции $t, t^2, \dots, t^n$ (функция $t^0 = 1$ уже использована для выявления веса $A = \frac{2}{n}$). В любом случае

приходим к нелинейной системе

$$\left\{ \begin{array}{l} t_1 + t_2 + \dots + t_n = 0, \\ t_1^2 + t_2^2 + \dots + t_n^2 = \frac{n}{3}, \\ t_1^3 + t_2^3 + \dots + t_n^3 = 0, \\ t_1^4 + t_2^4 + \dots + t_n^4 = \frac{n}{5}, \\ \dots\dots\dots \\ t_1^n + t_2^n + \dots + t_n^n = \frac{1+(-1)^n}{2} \cdot \frac{n}{n+1}. \end{array} \right. \quad (5.47)$$

Доказано, что эта специфическая симметричная система n уравнений с n неизвестными определяет единственный набор узлов $\{t_i\}_{i=1}^n$ при $n = 1, 2, 3, \dots, 7, 9$ (Чебышев) и не имеет действительных решений при $n = 8$ и $n \geq 10$ (Бернштейн^{*)}). Эти узлы подсчитаны с высокой точностью и затабулированы (при нескольких первых значениях n см. их далее в табл.5.3).

Вернувшись к более общему, чем (5.45), виду квадратурной формулы (5.44), замечаем, что эта формула имеет $2n$ параметров: n узлов t_i и n весов A_i . Если считать, что мы свободны в выборе как узлов, так и весов, можно попытаться подобрать их такими, чтобы равенство

$$\int_{-1}^1 \varphi(t) dt \approx \sum_{i=1}^n A_i \varphi(t_i) \quad (5.48)$$

было точным для многочленов степени $2n-1$ или, что то же, для $2n$ степенных функций $\varphi(t) = 1, t, t^2, \dots, t^{2n-1}$. В этом случае построение квадратурной формулы наивысшего алгебраического порядка точности вида (5.48), называемой **квадратурной формулой Гаусса**, упирается в решение нелинейной системы

$$\left\{ \begin{array}{l} \sum A_i = 2, \quad \sum A_i t_i = 0, \\ \sum A_i t_i^2 = \frac{2}{3}, \quad \sum A_i t_i^3 = 0, \\ \dots\dots\dots \\ \sum A_i t_i^{2n-2} = \frac{2}{2n-1}, \quad \sum A_i t_i^{2n-1} = 0, \end{array} \right. \quad (5.49)$$

^{*)} Бернштейн Сергей Натанович (1880–1968) — российский математик, академик АН СССР, один из основателей конструктивной теории функций.

(где i всюду изменяется от 1 до n), получающейся аналогично системе (5.47). Решение системы (5.49) весьма затруднительно, но его несложно обойти, если знать конечный результат.

Оказывается, узлами t_i квадратурной формулы Гаусса (5.48) служат корни многочлена Лежандра $\chi_n(t)$ (см. § 3.4), которые, как известно, существуют при любом n , различны и принадлежат интервалу $(-1, 1)$, а веса A_i находятся интегрированием базисных многочленов Лагранжа $l_i(t)$ степени $n-1$ (см. § 1.2), а именно:

$$A_i = \int_{-1}^1 \frac{(t-t_1)\dots(t-t_{i-1})(t-t_{i+1})\dots(t-t_n)}{(t_i-t_1)\dots(t_i-t_{i-1})(t_i-t_{i+1})\dots(t_i-t_n)} dt. \quad (5.50)$$

Докажем, что при таких t_i и A_i формула (5.48) будет точна при подстановке в нее вместо $\varphi(t)$ любого многочлена $P_{2n-1}(t)$ степени $2n-1$.

Согласно теореме о делении многочлена с остатком, для пары многочленов $P_{2n-1}(t)$ и $\chi_n(t)$ однозначно найдется другая пара многочленов $U_{n-1}(t)$ и $V_{n-1}(t)$ таких, что

$$P_{2n-1}(t) = U_{n-1}(t)\chi_n(t) + V_{n-1}(t). \quad (5.51)$$

В силу линейной независимости и попарной ортогональности (см. (3.24)) системы многочленов Лежандра $\{\chi_j\}_{j=0}^n$, найдется такой набор коэффициентов $\alpha_0, \alpha_1, \dots, \alpha_{n-1}$, что

$$U_{n-1}(t) = \alpha_0\chi_0(t) + \alpha_1\chi_1(t) + \dots + \alpha_{n-1}\chi_{n-1}(t) \quad \forall t \in [-1, 1]$$

и

$$\begin{aligned} \int_{-1}^1 U_{n-1}(t)\chi_n(t)dt &= \alpha_0 \int_{-1}^1 \chi_0(t)\chi_n(t)dt + \alpha_1 \int_{-1}^1 \chi_1(t)\chi_n(t)dt + \dots \\ &\dots + \alpha_{n-1} \int_{-1}^1 \chi_{n-1}(t)\chi_n(t)dt = 0 \quad \forall \alpha_0, \alpha_1, \dots, \alpha_{n-1}. \end{aligned}$$

Следовательно, интегрирование (5.51) приводит к равенству

$$\int_{-1}^1 P_{2n-1}(t)dt = \int_{-1}^1 V_{n-1}(t)dt,$$

являющемуся результатом преобразования левой части формулы (5.48) при $\varphi(t) = P_{2n-1}(t)$.

Рассмотрим теперь правую часть формулы (5.48) с A_i , определяемыми посредством (5.50), и с t_i такими, что $\chi_n(t_i) = 0$. Имеем:

$$\begin{aligned} \sum_{i=1}^n A_i P_{2n-1}(t_i) &= \sum_{i=1}^n A_i U_{n-1}(t_i) \chi_n(t_i) + \sum_{i=1}^n A_i V_{n-1}(t_i) = \\ &= \sum_{i=1}^n \int_{-1}^1 \frac{(t-t_1)\dots(t-t_{i-1})(t-t_{i+1})\dots(t-t_n)}{(t_i-t_1)\dots(t_i-t_{i-1})(t_i-t_{i+1})\dots(t_i-t_n)} dt \cdot V_{n-1}(t_i) = \\ &= \int_{-1}^1 \left[\sum_{i=1}^n \frac{(t-t_1)\dots(t-t_{i-1})(t-t_{i+1})\dots(t-t_n)}{(t_i-t_1)\dots(t_i-t_{i-1})(t_i-t_{i+1})\dots(t_i-t_n)} \cdot V_{n-1}(t_i) \right] dt = \int_{-1}^1 V_{n-1}(t) dt. \end{aligned}$$

Последнее равенство в цепочке проведенных преобразований объясняется тем, что выражение в квадратных скобках под знаком интеграла есть интерполяционный многочлен Лагранжа $(n-1)$ -й степени, составленный по n значениям $V_{n-1}(t_i)$ ($i = 1, 2, \dots, n$) многочлена $(n-1)$ -й степени $V_{n-1}(t)$, т.е. этот многочлен однозначно (в силу единственности интерполяционного многочлена) восстанавливается интерполированием.

Итак, имеем точное равенство

$$\int_{-1}^1 P_{2n-1}(t) dt = \sum_{i=1}^n A_i P_{2n-1}(t_i)$$

при указанном фиксировании величин A_i и t_i , означающее, что эти величины могут служить весами и узлами квадратурной формулы Гаусса (5.48).

Запишем теперь общую формулу для квадратур Чебышева и Гаусса применительно к исходному интегралу по промежутку $[a, b]$ на основе преобразования (5.43):

$$\int_a^b f(x) dx \approx \frac{b-a}{2} \sum_{i=1}^n A_i f\left(\frac{a+b}{2} + \frac{b-a}{2} t_i\right). \quad (5.52)$$

Значения узлов t_i и весов A_i , округленные до шести знаков после запятой, при $n = 2, 3, 4, 5$ приведены в табл.5.3. При необходимости узнать t_i , A_i с большей точностью или при $n > 5$ следует обратиться к другим источникам (например, [11, 46, 54, 57]).

Таблица 5.3

Узлы и веса квадратурных формул Чебышева и Гаусса

n	i	Формула Чебышева		Формула Гаусса	
		t_i	$A_i \equiv A$	t_i	A_i
2	1; 2	∓ 0.577350	1	∓ 0.577350	1
3	1; 3	∓ 0.707107	$\frac{2}{3}$	∓ 0.774597	$\frac{5}{9}$
	2	0		0	$\frac{8}{9}$
4	1; 4	∓ 0.794654	$\frac{1}{2}$	∓ 0.861136	0.347855
	2; 3	∓ 0.187592	$\frac{2}{2}$	∓ 0.339981	0.652145
5	1; 5	∓ 0.832497	$\frac{2}{5}$	∓ 0.906180	0.236927
	2; 4	∓ 0.374541		∓ 0.538469	0.478629
	3	0		0	0.568889

Остаточный член квадратурной формулы Гаусса (5.52), как следовало ожидать, выражается через $(2n)$ -ю производную функции $f(x)$:

$$r_n^{\Gamma} := I - I^{\Gamma} = \frac{(b-a)^{2n+1}(n!)^4}{(2n+1)[(2n)!]^3} f^{(2n)}(\xi_{\Gamma}), \quad \xi_{\Gamma} \in (a, b); \quad (5.53)$$

его вывод можно найти, например, в [11]. Там же выводятся и остаточные члены $r_n^{\chi} := I - I^{\chi}$ квадратур Чебышева вида (5.46) для всех конкретных значений n , при которых они существуют. Для r_n^{χ} , как и для остаточных членов формул Ньютона–Котеса, характерно «перепрыгивание» через порядок входящих в их выражение производных:

$$r_2^{\chi} = \frac{1}{135} \varphi^{IV}(\xi_2), \quad r_3^{\chi} = \frac{1}{360} \varphi^{IV}(\xi_3),$$

$$r_4^{\chi} = \frac{2}{42525} \varphi^{VI}(\xi_4), \quad r_5^{\chi} = \frac{13}{544320} \varphi^{VI}(\xi_5)$$

и т.д., где ξ_n — некоторые фиксированные точки интервала $(-1, 1)$. Сравнение этих величин с соответствующими значениями r_n^{Γ} , поставляемыми формулой (5.53) при $a = -1$, $b = 1$, т.е. для

квадратурной формулы Гаусса (5.48), а именно, со значениями

$$r_2^I = \frac{1}{135} \varphi^{IV}(\tilde{\xi}_2), \quad r_3^I = \frac{1}{15750} \varphi^{VI}(\tilde{\xi}_3),$$

$$r_4^I = \frac{2}{3472875} \varphi^{VIII}(\tilde{\xi}_4), \quad r_5^I = \frac{13}{1237732650} \varphi^X(\tilde{\xi}_5),$$

говорит о существенно более быстром убывании ошибки формулы Гаусса при достаточной гладкости интегрируемой функции.

Проблема корректного использования формул Чебышева и Гаусса (5.52) — в трудности оценивания модулей производных высоких порядков (если они еще существуют!). Выход можно искать в построении составных формул подобно тому, как это делалось выше. К сожалению, здесь неприменим напрямую принцип Рунге, которым можно было бы руководствоваться при разбиении интеграла на части для получения его значения с наперед заданной точностью. Полезную роль на этом пути может сыграть изучение *квадратурных формул Маркова*^{*)} [6, 11, 67]. Алгебраическая точность таких формул ниже, чем у соответствующих формул Гаусса, но зато какие-то узлы могут произвольно фиксироваться. Например, в число узлов «наси́льно» могут быть включены одна или обе границы интегрирования; в последнем случае такая формула называется *квадратурой Лобатто*^{**) [6, 9]}.

Замечание 5.2. По тому, входят границы интегрирования в число узлов или нет, квадратурные формулы подразделяются на *формулы замкнутого и открытого типов* соответственно. Очевидно, формулы Чебышева и Гаусса относятся к формулам открытого типа, а формулы Ньютона–Котеса — замкнутого. Введение квадратурных формул Маркова позволяет «замкнуть» формулы Гаусса.

Замечание 5.3. Если на промежутке интегрирования $[a, b]$ доступны значения не только подынтегральной функции, но и некоторых ее производных, можно отказаться от требования несовпадения узлов интегрирования, и от вида квадратурной формулы (5.42) перейти к более общему виду формул численного интегрирования с кратными узлами [57, 70]. В основу построения конкретных квадратурных формул такого типа можно положить, например, подмену подынтегральной функции $f(x)$ ее интерполяционным многочленом Эрмита $H_n(x)$ (см. § 1.8). Добавление к значениям функции только первых производных в точках a и b приводит к

^{*)} Ма́рков Андрей Андреевич (1856–1922) — широко известный русский математик, академик Петербургской академии наук.

^{**) Лоба́тто Рехюэл (1797–1866) — нидерландский математик.}

простейшей **формуле Эйлера** (Эйлера–Маклорена) [39]

$$I \approx \frac{b-a}{2} [f(a) + f(b)] + \frac{(b-a)^2}{12} [f'(a) - f'(b)],$$

на базе которой при равномерном разбиении (5.4) получается **составная формула Эйлера**

$$I \approx I^{\mathcal{E}} := h \left(\frac{y_0 + y_n}{2} + y_1 + y_2 + \dots + y_{n-1} \right) + \frac{h^2}{12} (f'(a) - f'(b)). \quad (5.54)$$

Последняя незначительно усложняет составную формулу трапеций (5.28), но существенно повышает ее точность, о чем свидетельствует величина

ошибки $r^{\mathcal{E}} := I - I^{\mathcal{E}} \approx \frac{h^4}{720} \int_a^b f^{IV}(x) dx$ в сравнении с величиной r^T в (5.29).

Интересно отметить, что формула (5.54) может быть выведена из формулы Грегори (см. замечание 5.1), если запись последней оборвать на втором слагаемом $\frac{1}{12}(\Delta y_0 - \Delta y_{k-1})$, зафиксировать $k = n$ и воспользоваться обобщавшейся в первой главе связью между конечными разностями и производными: $\Delta y_0 \approx hf'(a)$, $\Delta y_{n-1} \approx hf'(b)$.

§ 5.7. ФОРМУЛЫ ГАУССА–КРИСТОФФЕЛЯ

Квадратурной формулой Гаусса–Кристоффеля^{*)} [35, 39] (или формулой **типа Гаусса** [67], или просто **Гаусса** [9, 11, 81]) называют формулу наивысшего алгебраического порядка точности вида

$$\int_a^b p(x) f(x) dx \approx \sum_{i=1}^n A_i f(x_i), \quad (5.55)$$

где границы интегрирования a и b могут быть как конечными, так и бесконечными, а **весовая функция** $p(x)$ должна удовлетворять нескольким условиям. А именно, функция $p(x)$ должна быть непрерывна и положительна на интервале (a, b) ; при $x = a$ и $x = b$ она может обращаться в нуль или в бесконечность, при

этом должен существовать $\int_a^b p(x) dx$. Очевидно, на конечном

промежутке $[a, b]$ всем этим требованиям удовлетворяет функция $p(x) \equiv 1$, с которой квадратурная формула Гаусса вида (5.42) (или вида (5.48) при $a = -1, b = 1$) является частным случаем формулы Гаусса–Кристоффеля (5.55).

^{*)} Кристоффель Элвин Бруно (1829–1900) — немецкий математик.

Обобщим знания о квадратурной формуле Гаусса, полученные в предыдущем параграфе, на формулу (5.55). При этом, естественно, будем пользоваться более широким понятием ортогональности с весом, согласно которому множество многочленов $\{Q_0(x), Q_1(x), \dots, Q_n(x), \dots\}$ образует систему ортогональных с весовой функцией $p(x)$ многочленов на (a, b) , если

$$\int_a^b p(x) Q_k(x) Q_j(x) dx = \begin{cases} 0 & \text{при } k \neq j, \\ \text{const} \neq 0 & \text{при } k = j. \end{cases} \quad (5.56)$$

Как уже упоминалось в § 3.4, многочлен n -й степени $Q_n(x)$ из такой системы имеет на промежутке ортогональности (a, b) ровно n вещественных корней (независимо от того, конечен этот промежуток или бесконечен). Они и принимаются за узлы квадратурной формулы Гаусса—Кристоффеля.

Теорема 5.1. *Квадратурная формула (5.55) точна для произвольного многочлена степени $2n-1$, если ее узлами x_i служат корни многочлена $Q_n(x)$ из семейства многочленов, ортогональных на промежутке интегрирования (a, b) с весом $p(x)$, а весовыми коэффициентами — числа*

$$A_i = \int_a^b \frac{(x - x_1) \dots (x - x_{i-1})(x - x_{i+1}) \dots (x - x_n)}{(x_i - x_1) \dots (x_i - x_{i-1})(x_i - x_{i+1}) \dots (x_i - x_n)} p(x) dx. \quad (5.57)$$

Доказательство проведем схематично, поскольку оно повторяет рассуждения, которые были положены в обоснование аналогичного результата для формулы Гаусса (5.48).

Возьмем произвольный многочлен степени $2n-1$ и представим его в виде

$$P_{2n-1}(x) = U_{n-1}(x) Q_n(x) + V_{n-1}(x) \quad (5.58)$$

(ср. с (5.51)). Разложим $U_{n-1}(x)$ по базису $Q_0(x), Q_1(x), \dots, Q_{n-1}(x)$ с коэффициентами $\alpha_0, \alpha_1, \dots, \alpha_{n-1}$ и подставим $P_{2n-1}(x)$ в качестве $f(x)$ в левую часть формулы (5.55). Тогда, в силу (5.56), будем иметь

$$\begin{aligned} \int_a^b p(x) P_{2n-1}(x) dx &= \alpha_0 \int_a^b p(x) Q_0(x) Q_n(x) dx + \\ &+ \alpha_1 \int_a^b p(x) Q_1(x) Q_n(x) dx + \dots + \alpha_{n-1} \int_a^b p(x) Q_{n-1}(x) Q_n(x) dx + \\ &+ \int_a^b p(x) V_{n-1}(x) dx = \int_a^b p(x) V_{n-1}(x) dx. \end{aligned}$$

Подстановка в правую часть (5.55) функции $f(x) = P_{2n-1}(x)$ в виде (5.58) дает тот же результат:

$$\begin{aligned} \sum_{i=1}^n A_i P_{2n-1}(x_i) &= \sum_{i=1}^n A_i U_{n-1}(x_i) Q_n(x_i) + \sum_{i=1}^n A_i V_{n-1}(x_i) = 0 + \\ &+ \int_a^b \left[\sum_{i=1}^n \frac{(x-x_1)\dots(x-x_{i-1})(x-x_{i+1})\dots(x-x_n)}{(x_i-x_1)\dots(x_i-x_{i-1})(x_i-x_{i+1})\dots(x_i-x_n)} \cdot V_{n-1}(x_i) \right] p(x) dx = \\ &= \int_a^b p(x) V_{n-1}(x) dx, \end{aligned}$$

так как по условию $Q_n(x_i) = 0$, а

$$\sum_{i=1}^n \frac{(x-x_1)\dots(x-x_{i-1})(x-x_{i+1})\dots(x-x_n)}{(x_i-x_1)\dots(x_i-x_{i-1})(x_i-x_{i+1})\dots(x_i-x_n)} \cdot V_{n-1}(x_i) = V_{n-1}(x),$$

согласно теории интерполирования.

Как видим, для алгебраической функции $P_{2n-1}(x)$ равенство (5.55) является точным. Теорема доказана.

Условия на узлы x_i и веса A_i фигурируют в теореме 5.1 в качестве достаточных условий для того, чтобы формула (5.55) являлась квадратурной формулой наивысшего алгебраического порядка точности. Доказано [54 и др.], что они являются и необходимыми.

Замечание 5.4. При любом расположении узлов x_i на $[a, b]$ формулу вида (5.55) называют *интерполяционной квадратурной формулой* [67] или *квадратурной формулой интерполяционного типа* [6, 81], если ее весовые коэффициенты задаются равенством (5.57). Таковыми являются не только рассматриваемые здесь формулы Гаусса–Кристоффеля, но и изученные ранее простейшие формулы прямоугольников, трапеций, Симпсона, а также формулы Чебышева. Интерполяционными квадратурами не будут формулы численного интегрирования, опирающиеся, например, на аппроксимацию подынтегральной функции многочленами наилучших среднеквадратических приближений (см. гл. 3), а не на интерполяцию.

Несколько конкретных квадратурных формул Гаусса–Кристоффеля можно получить на базе приведенных в § 3.4 классических ортогональных многочленов.

Положив в формуле (5.55) $a = -1$, $b = 1$, $p(x) = \frac{1}{\sqrt{1-x^2}}$ и

взяв в качестве узлов x_i корни многочлена Чебышева $T_n(x)$ (см. (2.3)), придем к формуле

$$\int_{-1}^1 \frac{f(x)}{\sqrt{1-x^2}} dx \approx \sum_{i=1}^n A_i f(x_i), \quad (5.59)$$

которую называют **квадратурной формулой Эрмита** [11, 85]*). Формула (5.59) имеет алгебраический порядок точности $2n-1$, когда ее коэффициенты A_i , согласно теореме 5.1, вычисляются по формуле 5.57. Учитывая краткую запись интерполяционного многочлена Лагранжа в виде (1.6'), формулу для вычисления весовых коэффициентов в данном случае можно представить так:

$$A_i = \int_{-1}^1 \frac{T_n(x)}{(x-x_i)T'_n(x_i)} \cdot \frac{dx}{\sqrt{1-x^2}}. \quad (5.60)$$

Знание вида и свойств многочлена Чебышева позволяет найти значения интеграла в выражении (5.60) при $i=1, 2, \dots, n$ и установить, что все коэффициенты в формуле (5.59) равны между собой [11, 54, 67]:

$$A_i = \frac{\pi}{n} \quad \forall i \in \{1, 2, \dots, n\}.$$

Таким образом, квадратурная формула Эрмита есть

$$\int_{-1}^1 \frac{f(x)dx}{\sqrt{1-x^2}} \approx \frac{\pi}{n} \sum_{i=1}^n f(x_i), \quad (5.61)$$

где $x_i = \cos \frac{2i-1}{2n} \pi$ — корни многочлена $T_n(x)$. Известна она и как **формула Мелера** [67].

Промежуточное положение между общей формулой Гаусса–Кристоффеля (5.55) и ее частными представителями, каковыми являются формула Гаусса и формула Эрмита, занимает квадратурная формула вида

$$\int_{-1}^1 (1-x)^\alpha (1+x)^\beta f(x) dx \approx \sum_{i=1}^n A_i f(x_i) \quad (\alpha, \beta > -1). \quad (5.62)$$

На самом деле (5.62) — это двупараметрическое семейство квадратурных формул, имеющих наивысший алгебраический порядок точности, когда за узлы x_i принимаются корни ортогональных с фигурирующей под знаком интеграла весовой функцией

* В [46] так называется другая формула, см. далее (5.64).

многочленов Якоби (упоминавшихся в § 3.4), а коэффициенты A_i вычисляются в соответствии с формулой (5.57). При некоторых конкретных значениях параметров α и β узлы x_i и веса A_i можно найти затабулированными в справочной литературе (например, [57], где вместо $[-1, 1]$ за основу принят промежуток $[0, 1]$; там же содержатся числовые данные и для других формул численного интегрирования).

В случае произвольного замкнутого промежутка интегрирования $[a, b]$ для того, чтобы воспользоваться квадратурными формулами типа формулы (5.62), достаточно свести его, как это делалось ранее в § 5.6, к стандартному промежутку $[-1, 1]$.

Если одна или обе границы интегрирования бесконечны, тогда на основе ортогональных многочленов Лагерра и Эрмита приходим соответственно к формулам

$$\int_0^{\infty} e^{-x} f(x) dx \approx \sum_{i=1}^n A_i f(x_i) \quad (5.63)$$

и

$$\int_{-\infty}^{+\infty} e^{-x^2} f(x) dx \approx \sum_{i=1}^n A_i f(x_i), \quad (5.64)$$

которые иногда называют также соответственно **квадратурными формулами Лагерра и Эрмита** [46]^{*)}. За узлы в них принимают корни многочленов Лагерра $L_n(x)$ в (5.63) и Эрмита $H_n(x)$ в (5.64) (см. § 3.4), а весовые коэффициенты A_i подсчитываются по формулам [54]:

$$A_i = \int_0^{\infty} \frac{L_n(x)}{(x - x_i)L'_n(x_i)} e^{-x} dx = \frac{(n!)^2}{x_i [L'_n(x_i)]^2} = \left[\frac{(n-1)!}{nL_{n-1}(x_i)} \right]^2 x_i$$

для квадратуры (5.63) и

$$A_i = \int_{-\infty}^{+\infty} \frac{H_n(x)}{(x - x_i)H'_n(x_i)} e^{-x^2} dx = \frac{2^{n+1} n! \sqrt{\pi}}{[H'_n(x_i)]^2} = \frac{2^{n-1} (n-1)! \sqrt{\pi}}{nH_{n-1}^2(x_i)}$$

^{*)} Называют их также **квадратурными формулами с весом Чебышева–Лагерра** и **Чебышева–Эрмита** соответственно [45]. Имеется и более общая **формула Чебышева–Лагерра** [57]

$$\int_0^{\infty} x^{\alpha} e^{-x} f(x) dx \approx \sum_{i=1}^n A_i f(x_i),$$

поглощающая (5.63) при $\alpha = 0$.

для (5.64) (точки x_i в этих формулах предполагаются уже фиксированными указанным образом)^{*)}.

Числовые данные об узлах и весах квадратур Лагерра и Эрмита для значений n от 2 до 5 представлены в таблице 5.4 [46, 57].

Остаточные члены формул (5.63) и (5.64) имеют соответственно вид [54, 57]

$$R_n^{\mathcal{L}} = \frac{(n!)^2}{(2n)!} f^{(2n)}(\xi_{\mathcal{L}}) \quad \text{и} \quad R_n^{\mathcal{E}} = \frac{n! \sqrt{\pi}}{2^n (2n)!} f^{(2n)}(\xi_{\mathcal{E}}),$$

где $\xi_{\mathcal{L}} \in (0, +\infty)$, $\xi_{\mathcal{E}} \in (-\infty, +\infty)$.

Т а б л и ц а 5.4

Узлы и веса квадратурных формул Лагерра (5.63) и Эрмита (5.64)

n	i	Формула Лагерра		Формула Эрмита	
		x_i	A_i	x_i	A_i
2	1	0.585786	0.853553	-0.707107	0.886227
	2	3.414214	0.146447	0.707107	0.886227
3	1	0.415775	0.711093	-1.224745	0.295409
	2	2.294280	0.278518	0	1.181636
	3	6.289945	0.0103893	1.224745	0.295409
4	1	0.322548	0.603154	-1.650680	0.0813128
	2	1.745761	0.357419	-0.524648	0.804914
	3	4.536620	0.0388879	0.524648	0.804914
	4	9.395071	0.000539295	1.650680	0.0813128
5	1	0.263560	0.521756	-2.020183	0.0199532
	2	1.413403	0.398667	-0.958572	0.393619
	3	3.596426	0.0759424	0	0.945309
	4	7.085810	0.00361176	0.958572	0.393619
	5	12.640801	0.00002337	2.020183	0.0199532

^{*)} Чтобы в представленных здесь интегральных выражениях A_i узнать обоснованную в теореме 5.1 общую формулу (5.57), достаточно вспомнить краткую запись интерполяционного многочлена Лагранжа в виде (1.6'). Использование многочлена Лагерра $L_n(x)$ в первом и Эрмита $H_n(x)$ во втором интегральных выражениях подчеркивает, что за узлы $x_1, \dots, x_n$ принимаются корни именно этих многочленов.

§ 5.8. ПРИЕМЫ ПРИБЛИЖЕННОГО ВЫЧИСЛЕНИЯ НЕСОБСТВЕННЫХ ИНТЕГРАЛОВ

В этом параграфе рассмотрим вычисление интегралов вида

$$A)^*) \int_0^{\infty} f(x) dx,$$

$$Б) \int_{-\infty}^{+\infty} f(x) dx$$

или

$$В) \int_a^b f(x) dx$$

при наличии бесконечных разрывов у функции $f(x)$ в последнем интеграле либо в точке $x = a$, либо в точке $x = b$, либо, в общем случае, в некоторой точке $x = c \in [a, b]$. Вычисляемые несобственные интегралы, разумеется, предполагаются сходящимися.

Одним из источников получения численных значений несобственных интегралов А)–В) являются рассмотренные в предыдущем параграфе квадратурные формулы из класса формул Гаусса–Кристоффеля. Для их применения нужно выделить под интегралом подходящую весовую функцию и воспользоваться соответствующей квадратурой.

Так, в случае А) можно записать

$$\int_0^{\infty} f(x) dx = \int_0^{\infty} e^{-x} e^x f(x) dx \approx \sum_{i=1}^n A_i e^{x_i} f(x_i), \quad (5.65)$$

т.е. применить квадратурную формулу Лагерра (5.63).

Аналогично, для интеграла Б) имеем

$$\int_{-\infty}^{+\infty} f(x) dx = \int_{-\infty}^{+\infty} e^{-x^2} e^{x^2} f(x) dx \approx \sum_{i=1}^n A_i e^{x_i^2} f(x_i),$$

где A_i и x_i должны соответствовать формуле Эрмита (5.64).

^{*)} Более общий случай интегрирования на промежутке $[a; +\infty)$ сводится к случаю А) или линейной заменой переменной, или представлением интеграла $\int_a^{\infty}$ в виде двух интегралов: $\int_a^0$ и $\int_0^{\infty}$, один из которых — определенный.

Пример 5.2. Точное значение несобственного интеграла $\int_0^{\infty} \ln \operatorname{th} \frac{x}{2} dx$

равно $-\frac{\pi^2}{4} \approx -2.4674$ [28].

Применение к этому интегралу формулы (5.65), т.е. приближенного равенства

$$\int_0^{\infty} \ln \operatorname{th} \frac{x}{2} dx \approx \sum_{i=1}^n A_i \varphi(x_i),$$

где $\varphi(x) := e^x \ln \operatorname{th} \frac{x}{2} = \left(\ln \frac{1 - e^{-x}}{1 + e^{-x}} \right) / e^{-x}$, с A_i и x_i из табл.5.4, дает:

при $n = 2$

$$A_1 \varphi(x_1) + A_2 \varphi(x_2) \approx 0.8536 \cdot (-2.2562) + 0.1464 \cdot (-2.0007) \approx -2.219;$$

при $n = 4$

$$\sum_{i=1}^4 A_i \varphi(x_i) \approx -(0.6032 \cdot 2.5311 + 0.3574 \cdot 2.0207 + 0.03889 \cdot 2.0001 + 0.0005393 \cdot 2.0004) \approx -2.328.$$

В промежуточных вычислениях здесь обращает на себя внимание незначительное изменение от узла к узлу абсолютных значений функции $\varphi(x)$; быстрое убывание модуля подынтегральной функции $f(x)$ принимает на себя весовая функция e^{-x} , что находит отражение в убывании весовых коэффициентов A_i с ростом i . Как отмечается в [57], убывание или ограниченность модуля функции $e^x f(x)$ в (5.65) ($\varphi(x)$ в данном примере) чревата медленной сходимостью квадратурного процесса Лагерра, так как эта функция должна аппроксимироваться многочленами, а у них модули растут при $x \rightarrow \infty$. Отсюда — целесообразность введения множителя x^α (см. сноску на с. 179), который за счет подбора параметра α позволит регулировать поведение интегрируемой с весом $x^\alpha e^{-x}$ функции.

Так, в данном примере при $n = 4$, $\alpha = -\frac{1}{2}$, т.е. с весом Чебышева-Лагерра $e^{-x}/\sqrt{x}$, нужно просуммировать значения функции

$$\psi(x) := \sqrt{x} e^x \ln \operatorname{th} \frac{x}{2} = \frac{\sqrt{x}}{e^{-x}} \ln \frac{1 - e^{-x}}{1 + e^{-x}}$$

в узлах $x_1 = 0.145304$, $x_2 = 1.339097$, $x_3 = 3.926964$ и $x_4 = 8.588636$ с весовыми коэффициентами $A_1 = 1.322294$, $A_2 = 0.415605$, $A_3 = 0.0341560$ и $A_4 = 0.000399208$ соответственно [57]. Вычислив значения $\psi(x_1) \approx -1.1566$, $\psi(x_2) \approx -1.7696$, $\psi(x_3) \approx -3.9638$, $\psi(x_4) \approx -5.8581$

(возрастающие по модулю!), получаем приближение к данному интегралу ≈ -2.403 , более точное, чем найденное выше при том же числе узлов $n=4$ и параметре $\alpha = 0$.

К вычислению интегралов с бесконечной границей можно применять различные формулы численного интегрирования, пользуясь равенством, определяющим несобственный интеграл:

$$\int_a^\infty f(x)dx := \lim_{b \rightarrow \infty} \int_a^b f(x)dx .$$

Оно позволяет считать, что для достаточно больших значений b

$$\int_a^\infty f(x)dx \approx \int_a^b f(x)dx, \tag{5.66}$$

и вычислять определенный интеграл $\int_a^b f(x)dx$ с помощью известных квадратурных правил. Предполагая исходный интеграл абсолютно сходящимся, величину абсолютной погрешности, т.е. $\left| \int_b^\infty f(x)dx \right|$, за счет увеличения b можно сделать сколь угодно малой. Реально это может быть достигнуто либо с помощью аналитического оценивания этой величины, либо таким устройством алгоритма вычисления данного несобственного интеграла на основе приближенного равенства (5.66), при котором граница b вычисляемого определенного интеграла постепенно перемещалась бы в процессе счета вправо по оси абсцисс до тех пор, пока величина интеграла не перестанет изменяться в требуемых десятичных знаках.

В случае несобственных интегралов типа В) без ограничения общности можно считать, что подынтегральная функция имеет особенность на границе промежутка интегрирования, т.е. если точкой c , где $f(x)$ обращается в бесконечность, окажется внутренняя точка интервала (a, b) , то данный интеграл можно

представить символически как $\int_a^c + \int_c^b$. Также без потери общно-

сти, как мы уже знаем, достаточно рассматривать $\int_{-1}^1 f(x)dx$. Но к таким интегралам, в которых подынтегральная функция имеет

особыми точками -1 и/или 1 , можно применить квадратурную формулу Эрмита (Мелера) (5.61) в виде

$$\int_{-1}^1 f(x) dx = \int_{-1}^1 \frac{\sqrt{1-x^2} f(x)}{\sqrt{1-x^2}} dx \approx \frac{\pi}{n} \sum_{i=1}^n \sqrt{1-x_i^2} f(x_i) \quad (5.67)$$

или более общую формулу (5.62), где параметры $\alpha > -1$, $\beta > -1$ желательно подобрать так, чтобы функция

$$\varphi(x) := (1-x)^{-\alpha} (1+x)^{-\beta} f(x)$$

была как можно более гладкой. Такой прием при вычислении несобственных интегралов называют **мультипликативным выделением особенностей** [11, 35, 39]. Существует несколько специальных квадратурных формул, позволяющих «загнать» в весовые функции различные типы особенностей: степенную, логарифмическую и др. [57].

Другой прием работы с интегралами типа В), предложенный Л.В.Канторовичем, называется **аддитивным выделением особенностей**. Здесь, вообще говоря, нет необходимости заранее сводить интегрирование к стандартному промежутку. Суть приема состоит в том, что данный несобственный интеграл представляется в следующем виде:

$$\int_a^b f(x) dx = \int_a^b \varphi(x) dx + \int_a^b [f(x) - \varphi(x)] dx, \quad (5.68)$$

где функция $\varphi(x)$ подбирается так, чтобы она имела на $[a, b]$ такую же особенность, как и $f(x)$, и значение $\int_a^b \varphi(x) dx$ находилось аналитическим путем, а функция $f(x) - \varphi(x)$ должна быть гладкой, т.е. интеграл $\int_a^b [f(x) - \varphi(x)] dx$ без проблем должен вычисляться обычными квадратурами.

Для функций определенного класса со степенными особенностями разработана технология подбора таких функций $\varphi(x)$, которые принимают на себя особенность не только функции $f(x)$, но и некоторого числа ее производных, что позволяет ко второму интегралу в (5.68) применить квадратурные формулы заданного порядка [11, 30, 35]. Покажем, как это можно сделать.

Пусть $c \in [a, b]$ — особая точка функции $f(x)$ такой, что

$$f(x) = \frac{g(x)}{(x-c)^\gamma},$$

где $\gamma \in (0, 1)$, а $g(x) \in C^{k+1}[a; b]$, и пусть для вычисления второго интеграла в (5.68) предполагается использовать квадратурную формулу, требующую от функции $f(x) - \varphi(x)$ непрерывность k -ой производной. Введем в рассмотрение функцию

$$g_{k+1}(x) := g(c) + g'(c)(x-c) + \dots + \frac{g^{(k+1)}(c)}{(k+1)!}(x-c)^{k+1}.$$

Так как эта функция есть отрезок разложения Тейлора функции $g(x)$ в окрестности особой точки $x=c$, то, полагая

$$\varphi(x) := \frac{g_{k+1}(x)}{(x-c)^\gamma}, \text{ будем иметь равенство}$$

$$f(x) - \varphi(x) = \frac{g(x) - g_{k+1}(x)}{(x-c)^\gamma} = o[(x-c)^k],$$

означающее непрерывную дифференцируемость функции $f(x) - \varphi(x)$ не менее k раз.

Имеются обобщения рассмотренного преобразования подынтегральной функции, например, учитывающие особенности логарифмического типа [11].

Пример 5.3. Дан интеграл $I := \int_0^1 \frac{e^x}{\sqrt{x}} dx$. Его подынтегральная

функция $f(x) := \frac{e^x}{\sqrt{x}}$ имеет степенную особенность $\left(\gamma = \frac{1}{2}\right)$ в точке $x=0$

и удовлетворяет требуемым выше условиям. Используя известное разложение

$$e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \dots,$$

положим $g_3(x) := 1 + x + \frac{x^2}{2} + \frac{x^3}{6}$. Тогда

$$\varphi(x) := \frac{g_3(x)}{x^{\frac{1}{2}}} = x^{-\frac{1}{2}} + x^{\frac{1}{2}} + \frac{1}{2}x^{\frac{3}{2}} + \frac{1}{6}x^{\frac{5}{2}},$$

а

$$f(x) - \varphi(x) = \frac{6e^x - (6 + 6x + 3x^2 + x^3)}{6\sqrt{x}} \left(= \frac{1}{4!}x^{\frac{7}{2}} + \frac{1}{5!}x^{\frac{9}{2}} + \dots \right). \quad (5.69)$$

Согласно (5.68), непосредственно находим значение

$$I_1 := \int_0^1 \varphi(x) dx = \left[2\sqrt{x} + \frac{2}{3}x\sqrt{x} + \frac{1}{5}x^2\sqrt{x} + \frac{1}{21}x^3\sqrt{x} \right]_0^1 = \\ = 2\frac{32}{35} (\approx 2.914286),$$

а к интегралу $I_2 := \int_0^1 [f(x) - \varphi(x)] dx$ с полным основанием можно приме-

нить, например, эффективный алгоритм прямоугольников-трапеций (см. § 5.5), поскольку функция $f(x) - \varphi(x)$ имеет, по меньшей мере, непрерывную вторую производную. При этом из последнего представления этой функции в (5.69) легко видеть, что $f(0) - \varphi(0) = 0$. Подсчет значения

$\int_0^1 [f(x) - \varphi(x)] dx$ указанным алгоритмом с заданной точностью $\varepsilon = 10^{-6}$

дает $I_2 = 0.011018$ (за 65 вычислений подынтегральной функции), что приводит к значению $I = I_1 + I_2 = 2.925304$.

Часто требуемый эффект аддитивного выделения особенности достигается интегрированием «по частям».

Так, в данном примере такой прием оказывается более простым:

$$\int_0^1 \frac{e^x}{\sqrt{x}} dx = 2 \int_0^1 e^x d\sqrt{x} = 2\sqrt{x}e^x \Big|_0^1 - 2 \int_0^1 \sqrt{x}e^x dx = 2e - 2I_3,$$

где $I_3 := \int_0^1 \sqrt{x}e^x dx$ — определенный интеграл с «хорошей» подынтегральной функцией. Применение того же алгоритма численного интегрирования с той же заданной точностью приводит к значению $I_3 = 1.255630$ (за 2049 вычислений функции). В итоге получаем то же приближенное значение $I = 2e - 2I_3 = 2.925304$.

Мультипликативным выделением особенности в этом примере можно воспользоваться напрямую, если применить специальную квадратурную формулу вида

$$\int_0^1 x^\alpha g(x) dx \approx \sum_{i=1}^n A_i g(x_i) \quad (5.70)$$

при $\alpha = -0.5$ [57]:

$$\int_0^1 \frac{e^x}{\sqrt{x}} dx \approx \sum_{i=1}^n A_i e^{x_i}.$$

Например, при $n = 4$ с помощью таблицы узлов и весов формулы (5.70) из [57] находим

$$I \approx 0.725368 \cdot e^{0.033648} + 0.627413 \cdot e^{0.276184} + 0.444762 \cdot e^{0.634677} + 0.202457 \cdot e^{0.922157} \approx 2.925302,$$

а при $n = 6$ имеем

$$I \approx 0.498294 \cdot e^{0.015683} + 0.466985 \cdot e^{0.135300} + 0.406335 \cdot e^{0.344942} + 0.320157 \cdot e^{0.592750} + 0.213879 \cdot e^{0.817428} + 0.0943507 \cdot e^{0.963461} \approx 2.925304.$$

Налицо эффективность использования специальной квадратуры.

Применяются и иные приемы вычисления несобственных интегралов [9, 11, 39, 57, 77 и др.]. Заметим, что иногда достаточно сделать удачную замену переменной, чтобы преобразовать несобственный интеграл к более подходящему для вычисления виду.

УПРАЖНЕНИЯ

5.1. Выведите формулы остаточных членов для квадратурных формул левых и правых прямоугольников.

5.2. Сколько требуется знать значений подынтегральной функции для подсчета интеграла $\int_1^2 \frac{\ln x}{x} dx$ по формуле трапеций с точностью $\varepsilon = 0.01$?

5.3. а) Докажите, что если стоящая под знаком интеграла $I := \int_a^b f(x) dx$ функция $f(x)$ имеет знакостоянную на (a, b) вторую производную, то справедлива оценка

$$\left| I - \frac{I^{\Pi} + I^T}{2} \right| \leq \frac{1}{2} |I^{\Pi} - I^T|,$$

где I^{Π} и I^T — приближения к I , получаемые по формулам прямоугольников (5.8) и трапеций (5.28).

Каков содержательный смысл доказанного неравенства?

б) Можно ли утверждать, что значение интеграла $\int_0^{\pi/3} \ln \cos x \, dx$ заключено между его приближенными значениями I^{Π} и I^T ?

5.4. Найдите значения коэффициентов квадратурной формулы

$$\int_{-h}^h f(x) dx \approx Af(-h) + Bf(0) + Cf(h),$$

считая ее точной для многочленов второй степени; убедитесь, что она совпадает с формулой Симпсона.

5.5. Определите, с какой точностью можно вычислить $\int_0^1 \sin(e^x) dx$, привлекая девять значений подынтегральной функции:

- а) по формуле прямоугольников;
- б) по формуле трапеций;
- в) по формуле Симпсона.

5.6. Покажите, что содержащаяся в выражении (5.8) при $n = 1$ простейшая формула средней точки является частным случаем квадратурной формулы Гаусса.

5.7. Непосредственным рассмотрением систем (5.47) и (5.49) убедитесь в совпадении формул Чебышева и Гаусса при $n = 2$ и найдите точные значения узлов t_1, t_2 .

5.8. Рассмотрите разные частные случаи формулы Гаусса–Кристоффеля при $n = 1$ (квадратуры с одним узлом).

5.9. Подсчитайте коэффициенты квадратурной формулы Эрмита (5.59) для случая двух узлов непосредственно по формуле (5.60) (иначе, обоснуйте квадратурную формулу (5.61) при $n = 2$).

5.10. Постройте простейшую квадратурную формулу не интерполяционного типа, заменяя подынтегральную функцию линейной функцией наилучшего среднеквадратического приближения, определяемой по трем равноотстоящим узлам (см. функцию $\varphi_i(x)$ при построении линейного фильтра в § 4.1). На основе полученной простейшей формулы запишите составную формулу такого типа.

ГЛАВА 6 || АППРОКСИМАЦИЯ ПРОИЗВОДНЫХ

Посредством дифференцирования интерполяционных формул выводятся формулы для приближенного вычисления производных. Остаточные члены полученных приближенных формул находятся с помощью формулы Тейлора и дифференцированием соответствующих остаточных членов интерполяционных формул. Особая роль придается простейшим аппроксимациям первого и второго порядков точности первой и второй производных в узлах сетки, что существенно используется в последующих главах. Обращается внимание на существование таких шагов аппроксимации производных по формулам различных порядков точности, при которых ограниченная точность вычисленных значений функции наименьшим образом влияет на точность результата.

§ 6.1. ВЫВОД ФОРМУЛ ЧИСЛЕННОГО ДИФФЕРЕНЦИРОВАНИЯ

Численное дифференцирование, т.е. нахождение значений производных заданной функции $y = f(x)$ в заданных точках x , в отличие от рассмотренного в предыдущей главе численного интегрирования, можно считать не столь актуальной задачей в связи с отсутствием принципиальных трудностей с аналитическим нахождением производных. Однако имеется ряд моментов, не позволяющих обходить эту задачу стороной. Это и типичное для прикладных задач незнание аналитического вида $f(x)$, и возможное сильное усложнение функции при ее аналитическом дифференцировании (что затрудняет нахождение ее значений с высокой точностью), и желательность получения значений производных с помощью однотипных вычислительных процессов без привлечения аналитических выкладок. Главным же для дальнейшего является потребность в простых формулах, с помощью которых производные в заданных точках можно аппроксимировать несколькими значениями функции (быть может неизвестной) в этих и близких к ним точках.

Источником формул численного дифференцирования, как и большинства квадратурных формул, является полиномиальная интерполяция.

Зная в точках $x_i = x_0 + ih$ ($i = 0, 1, \dots, n$) при некотором $h > 0$ значения $y_i = f(x_i)$ данной функции $y = f(x)$, можно найти конечные разности $\Delta^k y_i$ и записать для нее, например, первый интерполяционный многочлен Ньютона $P_n(x)$ (см. (1.25)).

Дифференцируя приближенное равенство $f(x) \approx P_n(x)$, будем строить формулы приближенного дифференцирования разной точности в зависимости от степени n используемого интерполяционного многочлена. Через вспомогательную переменную $q := \frac{x - x_0}{h}$ приближенное представление функции $f(x)$ по первой формуле Ньютона выглядит наиболее просто:

$$f(x) \approx P_n(x_0 + qh) = y_0 + q\Delta y_0 + \frac{q(q-1)}{2!} \Delta^2 y_0 + \frac{q(q-1)(q-2)}{3!} \Delta^3 y_0 + \dots + \frac{q(q-1)\dots(q-n+1)}{n!} \Delta^n y_0 \quad (6.1)$$

(см. (1.26)). Отсюда получаем **конечноразностную формулу численного дифференцирования**

$$f'(x) \approx q'_x [P_n(x_0 + qh)]'_q = \frac{1}{h} \left(y_0 + q\Delta y_0 + \frac{q^2 - q}{2} \Delta^2 y_0 + \frac{q^3 - 3q^2 + 2q}{6} \Delta^3 y_0 + \frac{q^4 - 6q^3 + 11q^2 - 6q}{24} \Delta^4 y_0 + \dots \right)'_q,$$

т.е.

$$f'(x) \approx \frac{1}{h} \left(\Delta y_0 + \frac{2q-1}{2} \Delta^2 y_0 + \frac{3q^2 - 6q + 2}{6} \Delta^3 y_0 + \frac{4q^3 - 18q^2 + 22q - 6}{24} \Delta^4 y_0 + \dots \right). \quad (6.2)$$

При использовании последнего равенства для приближенного вычисления производной функции $f(x)$ в заданной точке $\tilde{x}$ из некоторой окрестности точки x_0 следует найти соответствующее значение $\tilde{q}$ переменной $q = \frac{x - x_0}{h}$ и подставить его в формулу (6.2). Максимальный порядок конечных разностей в этой формуле при желании получить производную с наибольшей точностью определяется в конкретной ситуации в соответствии с приведенными в §§ 1.4, 1.6 соображениями о выборе подходящей степени интерполяционного многочлена.

Аналогично можно вывести ряд других формул численного дифференцирования на основе различных интерполяционных формул (см. гл.1), более эффективно «работающих» вблизи других узловых точек и в общем случае не обязательно равноот-

стоящих. Предоставим читателю сделать это самостоятельно [11 и др.] и вернемся к приближенной формуле (6.2). Рассмотрим несколько ее частных случаев, фиксируя степень n лежащего в ее основе интерполяционного многочлена (6.1), равной 1, 2, 3. Этим значениям n отвечают соответственно одно, два, три первых слагаемых в формуле (6.2). Таким образом, имеем:

на основе линейной интерполяции

$$f'(x) \approx \frac{\Delta y_0}{h} \quad \text{для } x \in (x_0 - \delta, x_1 + \delta), \quad (6.3)$$

на основе квадратичной интерполяции

$$f'(x) \approx \frac{1}{h} \left(\Delta y_0 + \frac{2q-1}{2} \Delta^2 y_0 \right) \quad \text{для } x \in (x_0 - \delta, x_2 + \delta), \quad (6.4)$$

на основе кубической интерполяции

$$f'(x) \approx \frac{1}{h} \left(\Delta y_0 + \frac{2q-1}{2} \Delta^2 y_0 + \frac{3q^2-6q+2}{6} \Delta^3 y_0 \right) \quad (6.5)$$

для $x \in (x_0 - \delta, x_3 + \delta),$

и т.д. (при некоторых $\delta > 0$, определяющих промежуток экстраполяции приемлемого качества соответствующей интерполяционной формулой).

Для дальнейшего особый интерес представляют частные случаи формул (6.3)–(6.5), связывающие приближенное значение производной функции $f(x)$ в узлах $x_0, x_1, \dots$ с узловыми значениями самой функции. Учитывая, что точкам x_0, x_1, x_2, x_3 соответствуют значения $q = 0, 1, 2, 3$, и раскрывая конечные разности через значения y_i ($i = 0, 1, 2, 3$), имеем:

при $n = 1$ из (6.3)

$$f'(x_0) \approx y'_0 := \frac{y_1 - y_0}{h}, \quad (6.6)$$

$$f'(x_1) \approx y'_1 := \frac{y_1 - y_0}{h}; \quad (6.7)$$

при $n = 2$ из (6.4)

$$f'(x_0) \approx y'_0 := \frac{1}{2h} (-3y_0 + 4y_1 - y_2), \quad (6.8)$$

$$f'(x_1) \approx y'_1 := \frac{1}{2h} (y_2 - y_0), \quad (6.9)$$

$$f'(x_2) \approx y'_2 := \frac{1}{2h} (y_0 - 4y_1 + 3y_2); \quad (6.10)$$

при $n = 3$ из (6.5)

$$f'(x_0) \approx y'_0 := \frac{1}{6h}(-11y_0 + 18y_1 - 9y_2 + 2y_3),$$

$$f'(x_1) \approx y'_1 := \frac{1}{6h}(-2y_0 - 3y_1 + 6y_2 - y_3),$$

$$f'(x_2) \approx y'_2 := \frac{1}{6h}(y_0 - 6y_1 + 3y_2 + 2y_3),$$

$$f'(x_3) \approx y'_3 := \frac{1}{6h}(-2y_0 + 9y_1 - 18y_2 + 11y_3).$$

В случае необходимости этот ряд формул можно продолжить с помощью общей формулы (6.2) или посмотреть в других источниках [11 и др.].

Повторное дифференцирование приближенного равенства (6.1), т.е. взятие производной по x от правой части формулы (6.2) с учетом $\frac{dq}{dx} = \frac{1}{h}$, приводит к конечноразностной формуле вычисления второй производной

$$f''(x) \approx \frac{1}{h^2} \left[\Delta^2 y_0 + (q-1)\Delta^3 y_0 + \frac{1}{12}(6q^2 - 18q + 11)\Delta^4 y_0 + \dots \right], \quad (6.11)$$

из которой таким же образом следует приближенная формула для третьей производной

$$f'''(x) \approx \frac{1}{h^3} \left[\Delta^3 y_0 + \left(q - \frac{3}{2} \right) \Delta^4 y_0 + \dots \right],$$

и т.д.

Наиболее важной в приложениях является простейшая аппроксимация второй производной с помощью постоянной $\frac{\Delta^2 y_0}{h^2}$ на промежутке $(x_0 - \delta, x_2 + \delta)$, получающаяся из (6.11) фиксированием только одного слагаемого (случай $n = 2$). В частности, в точке x_1 имеем приближенное равенство

$$f''(x_1) \approx y''_1 := \frac{y_0 - 2y_1 + y_2}{h^2}, \quad (6.12)$$

которое вместе с формулами (6.6)–(6.10) широко используется при построении конечноразностных методов решения краевых задач для обыкновенных дифференциальных уравнений второго порядка (см. гл. 10) и для уравнений в частных производных.

§ 6.2. ОСТАТОЧНЫЕ ЧЛЕНЫ ПРОСТЕЙШИХ ФОРМУЛ ЧИСЛЕННОГО ДИФФЕРЕНЦИРОВАНИЯ

Чтобы получить представление о точности простейших аппроксимаций значений производных в узловых точках, определяемых формулами (6.6)–(6.10), (6.12), будем предполагать, что данная функция $f(x)$ обладает достаточной для вывода остаточных членов гладкостью. Кроме того, проведем в указанных формулах смещение индексов, т.е. будем считать, что исходная информация о функции соответствует изображенной на рис. 6.1, и речь идет об аппроксимации производных в i -м узле x_i и/или в отстоящих от него на расстоянии h узлах x_{i-1} и x_{i+1} .

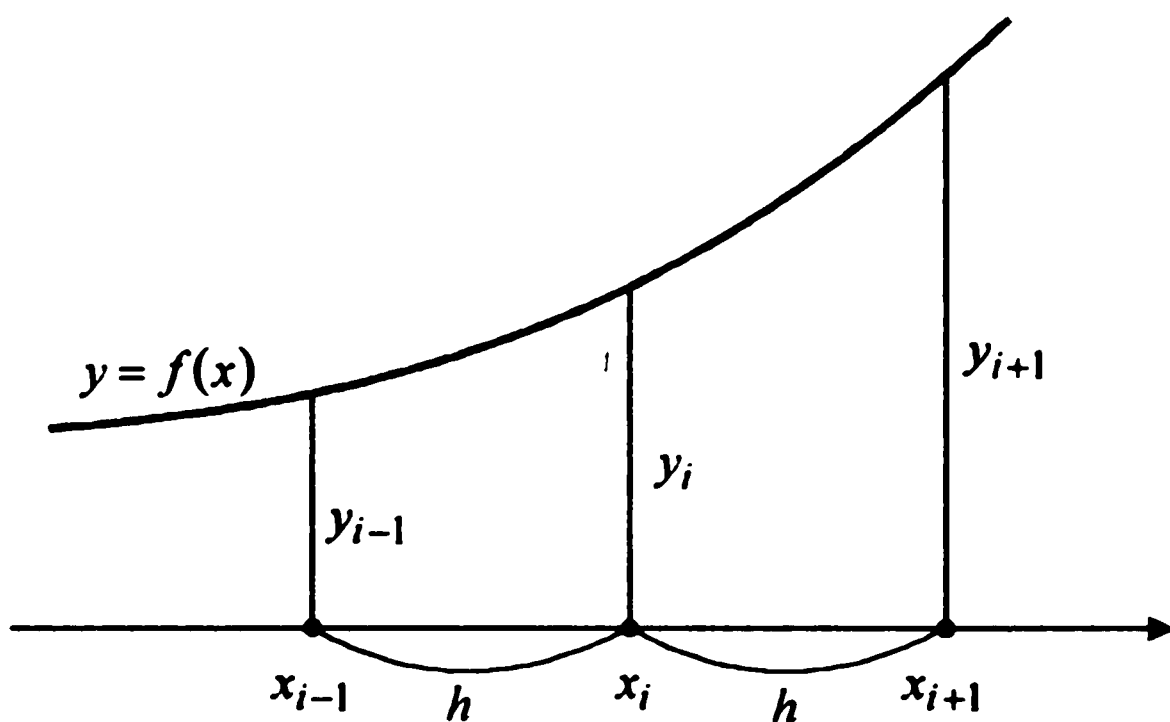


Рис. 6.1. К простейшей аппроксимации производных

Знание структуры приближенных выражений для производных, полученных из интерполяционных соображений, позволяет без особого труда (по крайней мере, для симметричных аппроксимаций) вывести формулы их остаточных членов, манипулируя разложениями $f(x)$ по формуле Тейлора подходящих порядков. Покажем это.

Простейшая несимметричная аппроксимация $f'(x_i)$ (формулы первого порядка точности). Запишем представление функции $f(x)$ по формуле Тейлора в окрестности точки x_i :

$$f(x) = f(x_i) + f'(x_i)(x - x_i) + \frac{f''(\xi)}{2}(x - x_i)^2.$$

Выразив отсюда $f'(x_i)$, имеем

$$f'(x_i) = \frac{f(x) - f(x_i)}{x - x_i} - \frac{f''(\xi)}{2}(x - x_i). \quad (6.13)$$

Первый член правой части этого равенства — разностное отношение, аппроксимирующее производную вблизи x_i , а второй — остаточный член, характеризующий точность такой аппроксимации. При фиксировании в (6.13) $x = x_{i-1}$ одновременно зафиксировается и неизвестная точка $\xi = \xi_{i-1} \in (x_{i-1}, x_i)$; таким образом, приходим к формуле левой аппроксимации $f'(x_i)$ с остаточным членом:

$$f'(x_i) = \frac{y_i - y_{i-1}}{h} + \frac{f''(\xi_{i-1})}{2}h. \quad (6.14)$$

Аналогично при $x = x_{i+1}$ из (6.13) получаем формулу правой аппроксимации $f'(x_i)$ с остаточным членом:

$$f'(x_i) = \frac{y_{i+1} - y_i}{h} - \frac{f''(\xi_{i+1})}{2}h. \quad (6.15)$$

В приближенных равенствах

$$f'(x_i) \approx \frac{y_{i+1} - y_i}{h} \quad (6.16)$$

при $i = 0$ и

$$f'(x_i) \approx \frac{y_i - y_{i-1}}{h} \quad (6.17)$$

при $i = 1$ узнаём выведенные ранее формулы (6.6), (6.7), а остаточные члены в (6.14); (6.15) говорят о том, что, пользуясь аппроксимациями (6.16), (6.17), мы совершаем ошибку $O(h)$, т.е. эти формулы имеют первый порядок точности. Определенную информацию об ошибках левой и правой аппроксимаций первого порядка дает знание знаков остаточных членов.

Простейшая симметричная аппроксимация $f'(x_i)$ (формула второго порядка точности). Из разложения

$$f(x) = f(x_i) + f'(x_i)(x - x_i) + \frac{1}{2}f''(x_i)(x - x_i)^2 + \frac{1}{6}f'''(\xi)(x - x_i)^3$$

при $x = x_{i+1}$ и $x = x_{i-1}$ имеем соответственно

$$f(x_{i+1}) = f(x_i) + f'(x_i)h + \frac{1}{2}f''(x_i)h^2 + \frac{1}{6}f'''(\xi_{i+1})h^3$$

и

$$f(x_{i-1}) = f(x_i) - f'(x_i)h + \frac{1}{2}f''(x_i)h^2 - \frac{1}{6}f'''(\xi_{i-1})h^3.$$

Выполнив почленное вычитание двух последних равенств, получаем

$$y_{i+1} - y_{i-1} = 2hf'(x_i) + \frac{h^3}{6}[f'''(\xi_{i+1}) + f'''(\xi_{i-1})],$$

откуда с помощью теоремы о среднем, примененной к сумме третьих производных в квадратных скобках, приходим к формуле симметричной аппроксимации $f'(x_i)$ с остаточным членом:

$$f'(x_i) = \frac{y_{i+1} - y_{i-1}}{2h} - \frac{h^2}{6} f'''(\xi_i), \quad (6.18)$$

где ξ_i — некоторая точка интервала (x_{i-1}, x_{i+1}) .

«Основная» часть формулы (6.18) —

$$f'(x_i) \approx \frac{y_{i+1} - y_{i-1}}{2h} \quad (6.19)$$

— при $i=1$ совпадает с (6.9), а вид ее остаточного члена $-\frac{h^2}{6} f'''(\xi_i)$ означает, что аппроксимация (6.19) имеет второй порядок точности относительно шага h .

Простейшие аппроксимации второй производной. Из представления

$$f(x) = f(x_i) + f'(x_i)(x - x_i) + \frac{f''(x_i)}{2}(x - x_i)^2 + \\ + \frac{f'''(x_i)}{6}(x - x_i)^3 + \frac{f^{IV}(\xi)}{24}(x - x_i)^4$$

имеем

$$f(x_{i+1}) = f(x_i) + hf'(x_i) + \frac{h^2 f''(x_i)}{2} + \frac{h^3 f'''(x_i)}{6} + \frac{h^4 f^{IV}(\xi_{i+1})}{24}, \\ f(x_{i-1}) = f(x_i) - hf'(x_i) + \frac{h^2 f''(x_i)}{2} - \frac{h^3 f'''(x_i)}{6} + \frac{h^4 f^{IV}(\xi_{i-1})}{24},$$

откуда почленным сложением получаем

$$y_{i+1} + y_{i-1} = 2y_i + h^2 f''(x_i) + \frac{h^4}{24}[f^{IV}(\xi_{i+1}) + f^{IV}(\xi_{i-1})].$$

Выражая из последнего равенства $f''(x_i)$, приходим к формуле симметричной аппроксимации $f''(x_i)$ с остаточным членом:

$$f''(x_i) = \frac{y_{i-1} - 2y_i + y_{i+1}}{h^2} - \frac{h^2}{12} f^{IV}(\xi_i). \quad (6.20)$$

Остаточный член этой формулы $-\frac{h^2}{12}f^{IV}(\xi_i)$ с некоторым $\xi_i \in (x_{i-1}, x_{i+1})$ характеризует приближенное равенство

$$f''(x_i) \approx \frac{y_{i-1} - 2y_i + y_{i+1}}{h^2} \quad (6.21)$$

как аппроксимацию второй производной в точке x_i второго порядка точности, т.е. с погрешностью $O(h^2)$.

То же отношение $\frac{y_{i-1} - 2y_i + y_{i+1}}{h^2}$, используемое в качестве несимметричной аппроксимации второй производной функции $f(x)$, т.е. для вычисления приближенных значений $f''(x_{i-1})$ и $f''(x_{i+1})$, дает лишь первый порядок точности. Действительно, по формуле Тейлора для второй производной имеем

$$f''(x) = f''(x_i) + f'''(\xi)(x - x_i).$$

Подставляя сюда вместо $f''(x_i)$ правую часть равенства (6.20), получаем

$$f''(x) = \frac{y_{i-1} - 2y_i + y_{i+1}}{h^2} + f'''(\xi)(x - x_i) - \frac{h^2}{12}f^{IV}(\xi_i).$$

Из этого равенства при $x = x_{i+1}$ и $x = x_{i-1}$ следуют частные формулы несимметричной аппроксимации второй производной с остаточными членами:

$$f''(x_{i-1}) = \frac{y_{i-1} - 2y_i + y_{i+1}}{h^2} - hf'''(\xi_{i-1}) - \frac{h^2}{12}f^{IV}(\xi_i)$$

и

$$f''(x_{i+1}) = \frac{y_{i-1} - 2y_i + y_{i+1}}{h^2} + hf'''(\xi_{i+1}) - \frac{h^2}{12}f^{IV}(\xi_i).$$

С помощью формулы Тейлора можно вывести остаточные члены и других простейших аппроксимаций производных.

Более общий подход для получения выражений остаточных членов интерполяционных формул численного дифференцирования состоит в дифференцировании остаточного члена интерполяционной формулы Лагранжа [11, 30]. Согласно (1.13),

$$R_n(x) := f(x) - L_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \Pi_{n+1}(x), \quad (6.22)$$

где $L_n(x)$ — интерполяционный многочлен Лагранжа, построенный для $n+1$ раз дифференцируемой функции $f(x)$ по $n+1$ узлу $x_0, x_1, \dots, x_n$ (неважно, в какой форме), ξ — некоторая точка из интервала (x_0, x_n) , а $\Pi_{n+1}(x) := (x - x_0)(x - x_1) \dots (x - x_n)$. Из (6.22) следует

$$R'_n(x) := f'(x) - L'_n(x) = \frac{1}{(n+1)!} \left\{ \frac{d}{dx} [f^{(n+1)}(\xi)] \Pi_{n+1}(x) + f^{(n+1)}(\xi) \frac{d}{dx} [\Pi_{n+1}(x)] \right\}. \quad (6.23)$$

Если величина $\frac{d}{dx} [f^{(n+1)}(\xi)]$ ограничена, то при подстановке в последнее выражение узловых значений $x = x_i$ ($i = 0, 1, \dots, n$) за счет $\Pi_{n+1}(x_i) = 0$ получим простую формулу остаточного члена аппроксимаций $f'(x_i) \approx L'(x_i)$ первой производной в узлах интерполяции:

$$f'(x_i) - L'(x_i) = \frac{f^{(n+1)}(\xi_i)}{(n+1)!} \Pi'_{n+1}(x_i), \quad \xi_i \in (x_0, x_n). \quad (6.24)$$

В случае равноотстоящих узлов $x_i = x_0 + ih$, который здесь, в основном, и рассматривается,

$$\Pi'_{n+1}(x_i) = (-1)^{n-i} i! (n-i)! h^n,$$

вследствие чего равенство (6.24) трансформируется в формулу

$$f'(x_i) - y'_i = (-1)^{n-i} \frac{i!(n-i)!}{(n+1)!} h^n f^{(n+1)}(\xi_i). \quad (6.25)$$

Из нее отчетливо видно, что при аппроксимации первой производной в точках x_i значениями $y'_i := L'_n(x_i)$, получаемыми дифференцированием интерполяционного многочлена n -й степени, остаточный член имеет n -й порядок относительно шага аппроксимации h .

Воспользуемся формулой (6.25), чтобы выписать остаточные члены несимметричных аппроксимаций первой производной (6.8) и (6.10). При $n = 2$, $i = 0$ и $i = 2$ из (6.25) следует

$$f'(x_0) - y'_0 = \frac{h^2}{3} f'''(\xi_0)$$

и

$$f'(x_2) - y'_2 = \frac{h^2}{3} f'''(\xi_2).$$

Ставя эти формулы в один ряд с формулой (6.18), имеем:

$$f'(x_{i-1}) = \frac{-3y_{i-1} + 4y_i - y_{i+1}}{2h} + \frac{h^2}{3} f'''(\xi_{i-1}), \quad (6.26)$$

$$f'(x_{i+1}) = \frac{y_{i-1} - 4y_i + 3y_{i+1}}{2h} + \frac{h^2}{3} f'''(\xi_{i+1}). \quad (6.27)$$

Заметим, что формулы несимметричной аппроксимации $f'(x_{i-1})$ (6.26) и $f'(x_{i+1})$ (6.27) второго порядка точности имеют в остаточном члене вдвое больший коэффициент, чем формула симметричной аппроксимации (6.18).

Для оценивания погрешности численного дифференцирования при значениях аргумента, не совпадающих с узловыми, и для получения остаточных членов приближенных формул $f^{(k)}(x) \approx L_n^{(k)}(x)$ при $k > 1$ формула (6.23) малоприспособна. В книге [11] на основе интерполяционной формулы Ньютона для неравных промежутков (1.43) и связей между разделенными разностями и производными выводятся следующие формулы остаточных членов:

$$f'(x) - L'_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \Pi'_{n+1}(x) + \frac{f^{(n+2)}(\xi_1)}{(n+2)!} \Pi_{n+1}(x), \quad (6.28)$$

$$\begin{aligned} f''(x) - L''_n(x) = & \frac{f^{(n+1)}(\xi)}{(n+1)!} \Pi''_{n+1}(x) + \\ & + 2 \frac{f^{(n+2)}(\xi_1)}{(n+2)!} \Pi'_{n+1}(x) + 2 \frac{f^{(n+3)}(\xi_2)}{(n+3)!} \Pi_{n+1}(x), \end{aligned} \quad (6.29)$$

и вообще, для $k \in \{0, 1, \dots, n\}$

$$\begin{aligned} f^{(k)}(x) - L_n^{(k)}(x) = \\ = \sum_{j=0}^k \frac{k!}{(k-j)!(n+j+1)!} f^{(n+j+1)}(\xi_j) \Pi_{n+1}^{(k-j)}(x). \end{aligned} \quad (6.30)$$

Имеются и другие, отличные от (6.28)–(6.30) представления остаточных членов и оценок погрешностей интерполяционных формул численного дифференцирования [21, 56 и др.].

Так как остаточные члены формул численного дифференцирования выражаются через производные более высоких порядков, чем искомые производные, то «в чистом виде» подобные выражения остаточных членов малоприспособны для практического оценивания погрешностей. Однако не стоит считать их совсем

бесполезными. Во-первых, они важны для качественного сравнения (например, по порядку) различных аппроксимаций производных. Во-вторых, ими можно воспользоваться для оценивания точности результатов численного дифференцирования, пользуясь простейшими связями между производными и конечными (или разделенными) разностями, например, составляя таблицы конечных разностей до требуемого в оценке порядка и принимая за $\max |f^{(n+1)}(x)|$ величину $\max \{ |\Delta^{n+1} y_i| \} / h^{n+1}$ в соответствующей области таблицы. В третьих, выполнение условий, при которых справедливы те или иные формулы аппроксимации производных на равномерной сетке, позволяет при вычислении значений производных применять принцип Рунге двойного счета с разными шагами подобно тому, как это делалось в § 5.5 при вычислении интегралов (более подробно об этом см. в [39, 56]).

Пример 6.1. Бесконечно гладкая функция $y = f(x)$ задана таблицей своих значений и значений конечных разностей (табл.6.1).

Таблица 6.1

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$	$\Delta^5 y$
1.0	0.367879					
		-0.035008				
1.1	0.332871		0.003331			
		-0.031677		-0.000316		
1.2	0.301194		0.003015		0.000028	
		-0.028662		-0.000288		0.000001
1.3	0.272532		0.002727		0.000029	
		-0.025935		-0.000259		-0.000004
1.4	0.246597		0.002468		0.000025	
		-0.023467		-0.000234		-0.000005
1.5	0.223130		0.002234		0.000020	
		-0.021233		-0.000214		0.000002
1.6	0.201897		0.002020		0.000022	
		-0.019213		-0.000192		-0.000003
1.7	0.182684		0.001828		0.000019	
		-0.017385		-0.000173		-0.000005
1.8	0.165299		0.001655		0.000014	
		-0.015730		-0.000159		
1.9	0.149569		0.001496			
		-0.014234				
2.0	0.135335					

Найдем приближения к значению $f'(1.5)$, пользуясь простейшими аппроксимациями по формулам первого и второго порядков точности. Имеем:

по формуле (6.16)

$$f'(1.5) \approx \frac{f(1.6) - f(1.5)}{0.1} \approx -0.21233;$$

по формуле (6.17)

$$f'(1.5) \approx \frac{f(1.5) - f(1.4)}{0.1} \approx -0.23467;$$

по формуле (6.19)

$$f'(1.5) \approx \frac{f(1.6) - f(1.4)}{0.2} \approx -0.22350; \quad (6.31)$$

по формуле (6.26)

$$f'(1.5) \approx \frac{-3f(1.5) + 4f(1.6) - f(1.7)}{0.2} \approx -0.22243;$$

по формуле (6.27)

$$f'(1.5) \approx \frac{f(1.3) - 4f(1.4) + 3f(1.5)}{0.2} \approx -0.22233.$$

Если на основании данной таблицы конечных разностей посчитать, что

$$\max_{x \in [1, 2]} |f''(x)| \approx \frac{\max |\Delta^2 y|}{h^2} \approx 0.3 \quad \text{и} \quad \max_{x \in [1, 2]} |f'''(x)| \approx \frac{\max |\Delta^3 y|}{h^3} \approx 0.3,$$

то в соответствии с выведенными выше формулами остаточных членов находим следующие оценки абсолютных погрешностей для полученных приближенных значений $f'(1.5)$:

$$\begin{aligned} &\approx \frac{0.3}{2} \cdot 0.1 = 0.015 \quad \text{для первых двух,} \\ &\approx \frac{0.3}{6} \cdot 0.01 = 0.0005 \quad \text{для третьего,} \\ &\approx \frac{0.3}{3} \cdot 0.01 = 0.001 \quad \text{для двух последних.} \end{aligned}$$

Откроем теперь факт, что исходная таблица составлена для функции e^{-x} и, значит, $f'(x) = -f(x)$. Это позволяет сравнить полученные приближенные значения $f'(1.5)$ с истинным $f'(1.5) = -f(1.5)$. Разница между приближенными значениями $f'(1.5)$ (без учета знака) составляет соответственно

$$0.01080, \quad 0.01114, \quad 0.00037, \quad 0.00070, \quad 0.00080$$

— числа, достаточно хорошо вписывающиеся в подсчитанные оценки

абсолютных погрешностей *).

Попытаемся при вычислении $f'(1.5)$ применить принцип Рунге. Считая, что невозможно пополнить таблицу новыми значениями функции, остается заложить процесс увеличения шага, т.е. включить механизм прореживания таблицы.

Примем за основу формулу симметричной аппроксимации второго порядка точности, а за начальный шаг — шаг $h = 0.1$ исходной таблицы. Обозначив приближенное значение $f'(1.5)$ через $y'(h)$, в соответствии с правилом Рунге уточняем его по формуле

$$y'(h) + \frac{y'(h) - y'(2h)}{3}.$$

Учитывая, что $y'(h) \approx -0.22350$ (см. (6.31)), а

$$y'(2h) = \frac{f(1.7) - f(1.3)}{0.4} \approx -0.22462,$$

находим поправку Ричардсона

$$\frac{y'(h) - y'(2h)}{3} \approx 0.000373,$$

прибавление которой к $y'(h)$ дает значение $f'(1.5) \approx -0.223127$, лишь в последнем знаке отличающееся от точного. Модуль этой поправки служит приближенной оценкой погрешности значения $y'(h)$, как видим, неплохо согласующейся по порядку с полученной выше оценкой (0.0005).

Для $f''(1.5)$ по симметричной формуле (6.21) с шагом $h = 0.1$ имеем

$$y''(0.1) = \frac{f(1.4) - 2f(1.5) + f(1.6)}{0.01} \approx 0.2234,$$

с шагом $2h = 0.2$ —

$$y''(0.2) = \frac{f(1.3) - 2f(1.5) + f(1.7)}{0.04} \approx 0.2239.$$

Поправка Ричардсона равна

$$\frac{0.2234 - 0.2239}{3} \approx -0.00017;$$

ее прибавление к значению $y''(0.1)$ дает уточненное значение

$$f''(1.5) \approx 0.2234 - 0.00017 = 0.22323.$$

В соответствии с выражением остаточного члена в (6.20), принимая

$$\max_{x \in [1, 2]} |f'''(x)| \approx \frac{\max |\Delta^4 y|}{h^4} \approx 0.3,$$

* Указанные оценки могут быть немного точнее, если максимумы модулей конечных разностей при оценивании модулей производных, входящих в остаточные члены, учитывать только в используемой части таблицы.

находим оценку погрешности значения $y''(0.1)$:

$$|f''(1.5) - y''(0.1)| \leq \frac{0.3}{12} \cdot 0.01 = 0.00025;$$

близок к ней и модуль поправки Ричардсона. Однако истинная ошибка $0.2234 - 0.22313 = 0.00027$, хотя и имеет тот же порядок, но превосходит ту и другую оценки. Это можно объяснить влиянием погрешностей округления исходных данных, более заметным при приближенном вычислении старших производных (см. следующий параграф). Нетрудно видеть, что по содержащимся в таблице данным найти какие-либо значения пятой и последующих производных в принципе невозможно.

Замечание 6.1. Обратим внимание на одинаковость структуры остаточных членов формул численного дифференцирования и интегрирования, опирающихся на полиномиальное интерполирование данной функции по системе равноотстоящих узлов с шагом h . Вхождение в выражения остаточных членов старших производных, порядок которых определенным образом согласован с показателем степени шага, наводит на мысль о возможности и здесь, т.е. при численном дифференцировании, пользоваться понятием алгебраического порядка точности. Структура же самих формул приближенного дифференцирования, имеющих вид

$$f^{(k)}(x) \approx \sum_{i=1}^n c_i f(x_i), \quad k \in \mathbb{N}, \quad (6.32)$$

позволяет строить конкретные формулы такого типа методом неопределенных коэффициентов, подбирая коэффициенты c_i в (6.32) (и, возможно, узлы x_i) так, чтобы формула (6.32) была точна для произвольных многочленов некоторой фиксированной степени (например, наиболее высокой при заданном h) [8].

Так, желая получить приближенную формулу

$$f''(x) \approx c_1 f(x_i - h) + c_2 f(x_i) + c_3 f(x_i + h),$$

будем подставлять в нее вместо $f(x)$ последовательно степенные функции $1, x, x^2$, вторые производные которых при любом x равны соответственно $0, 0, 2$. Получаем систему

$$\begin{cases} c_1 + c_2 + c_3 = 0, \\ c_1(x_i - h) + c_2 x_i + c_3(x_i + h) = 0, \\ c_1(x_i - h)^2 + c_2 x_i^2 + c_3(x_i + h)^2 = 2. \end{cases}$$

Считая, что в качестве x_i может быть взята любая точка, полагаем $x_i = 0$, и из упрощенной таким образом системы

$$\begin{cases} c_1 + c_2 + c_3 = 0, \\ (c_3 - c_1)h = 0, \\ (c_1 + c_3)h^2 = 2 \end{cases}$$

находим $c_1 = c_3 = \frac{1}{h^2}$, $c_2 = -\frac{2}{h^2}$. В результате приходим к формуле

$$f''(x) = \frac{f(x_{i-1}) - 2f(x_i) + f(x_{i+1}))}{h^2},$$

при $x \in [x_{i-1}, x_{i+1}]$ точной для многочленов второй степени, а при $x = x_i$, как нам уже известно (см. (6.20)), она точна для многочленов третьей степени.

§ 6.3. ОПТИМИЗАЦИЯ ШАГА ЧИСЛЕННОГО ДИФФЕРЕНЦИРОВАНИЯ ПРИ ОГРАНИЧЕННОЙ ТОЧНОСТИ ЗНАЧЕНИЙ ФУНКЦИИ

Откажемся от использовавшегося ранее обозначения $y_i = f(x_i)$ и будем полагать, что y_i — это приближенное значение функции $f(x_i)$ в точке x_i . Предположим, что уровень абсолютных погрешностей значений y_i в разных узлах x_i примерно одинаков и ограничен числом $\delta > 0$. Таким образом, имеем

$$f(x_i) \approx y_i (\pm \delta), \quad \text{т.е. } f(x_i) \in (y_i - \delta, y_i + \delta).$$

Будем изучать влияние неточностей задания (вычисления) значений функции на результаты аппроксимации производных по этим значениям для различных простейших формул.

Рассмотрим простейшую формулу (6.15) аппроксимации первой производной в узле x_i правым разностным отношением:

$$f'(x_i) = \frac{f(x_{i+1}) - f(x_i)}{h} - \frac{h}{2} f''(\xi_{i+1}).$$

Так как при вычитании приближенных чисел их абсолютные погрешности складываются, то замена неизвестного точно разност-

ного отношения $\frac{f(x_{i+1}) - f(x_i)}{h}$ реально вычисляемым отноше-

нием $\frac{y_{i+1} - y_i}{h}$ порождает ошибку, оцениваемую по модулю ве-

личины $\frac{2\delta}{h}$. А поскольку погрешность аппроксимации $f'(x)$

точным отношением $\frac{f(x_{i+1}) - f(x_i)}{h}$ в соответствии с выражением остаточного члена можно оценить величиной $\frac{M_2}{2}h$, где

$M_2 := \max_{x \in [x_i, x_{i+1}]} |f''(x)|$, приходим к неравенству

$$\left| f'(x_i) - \frac{y_{i+1} - y_i}{h} \right| \leq \frac{2\delta}{h} + \frac{M_2}{2}h.$$

Итак, полная абсолютная погрешность приближенного равенства

$$f'(x_i) \approx \frac{y_{i+1} - y_i}{h} \quad (6.33)$$

оценивается величиной

$$g(h) := \frac{2\delta}{h} + \frac{M_2}{2}h, \quad (6.34)$$

образованной двумя слагаемыми: вычислительной погрешностью, порождаемой неточными значениями функции, и погрешностью аппроксимации, связанной с выбором формулы численного дифференцирования. При $h \rightarrow 0$ за счет первого слагаемого функция $g(h)$ бесконечно растет; при сравнительно больших h с ростом h также будет наблюдаться рост $g(h)$ (асимптотически линейный), благодаря второму слагаемому. Схематично поведение графика функции $g = g(h)$ отображено на рис. 6.2.

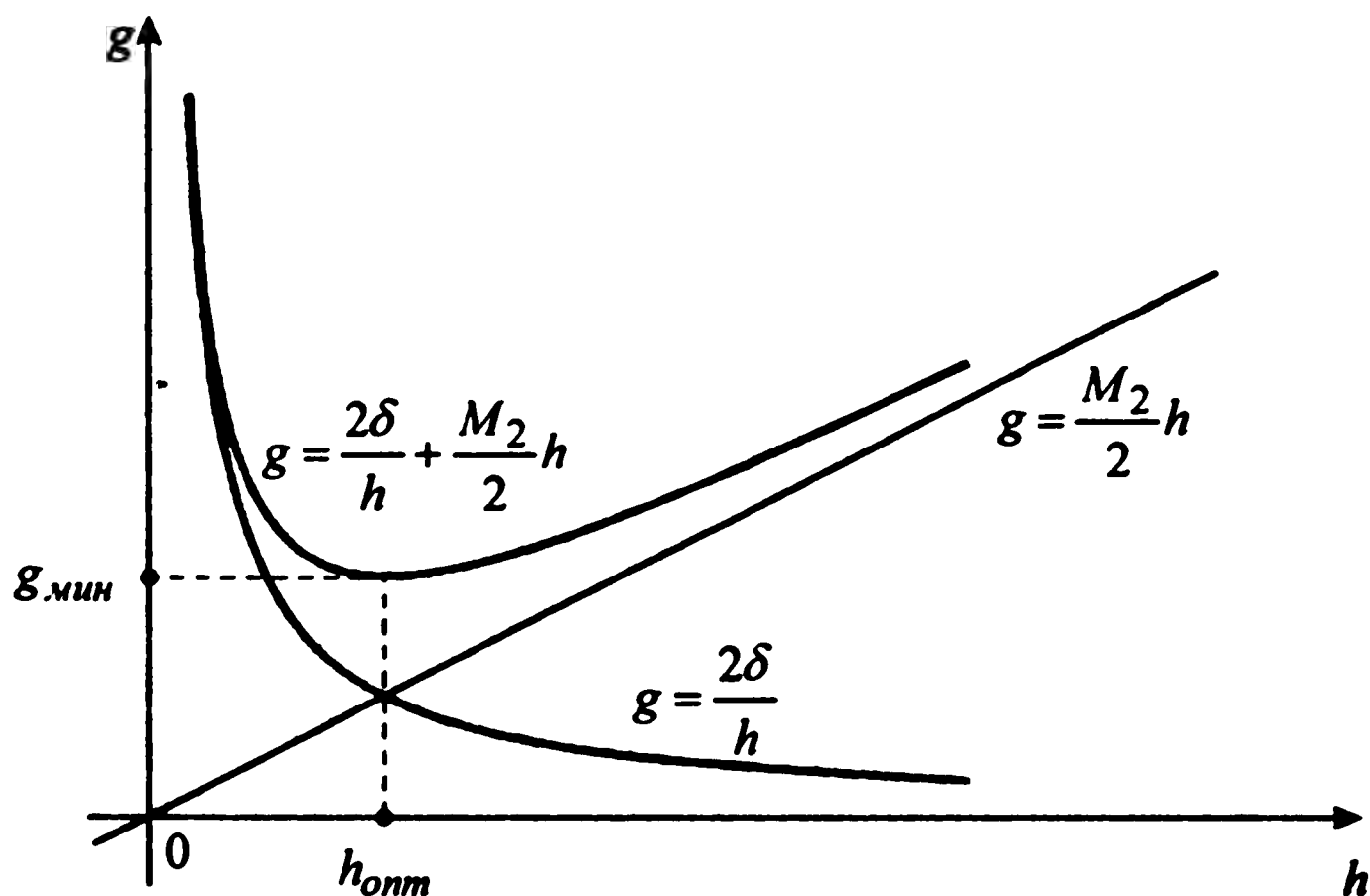


Рис. 6.2. Зависимость границы полной погрешности (6.34) приближенного равенства (6.33) от шага h

Не вызывает сомнений существование такого значения $h = h_{opt}$, при котором верхняя граница $g(h)$ полной погрешности приближенного равенства (6.33) минимальна. Это значение легко найти, приравняв нулю производную функции $g(h)$. Имеем:

$$-\frac{2\delta}{h^2} + \frac{M_2}{2} = 0,$$

откуда $h_{opt} = 2\sqrt{\frac{\delta}{M_2}}$ — *оптимальный шаг*. Из этой формулы

видно, что сделать аппроксимацию производной в точке x_i приближенным разностным отношением (6.33) сколь угодно качественной можно лишь в том случае, когда величину δ можно сделать сколь угодно малой (тогда можно рассчитывать на малость h_{opt} , что, в свою очередь, обеспечит малость $g_{min} = g(h_{opt})$). Реально же при машинных вычислениях дело обстоит следующим образом [8].

В лучшем случае, значения функции по точности ограничены лишь разрядностью мантиссы ЭВМ. Тогда, грубо полагая

$\delta = 2^{-m}$, $M_2 = 1$, находим, что $h_{opt} = 2 \cdot 2^{\frac{m}{2}}$ и, следовательно.

$$g_{min} = \frac{2 \cdot 2^{-m}}{2 \cdot 2^{\frac{m}{2}}} + \frac{2 \cdot 2^{\frac{m}{2}}}{2} = 2 \cdot 2^{\frac{m}{2}}. \quad (6.35)$$

Этот результат говорит о том, что *при простейших аппроксимациях производной с помощью правых разностных отношений (очевидно, для левых — то же самое) происходит потеря точности, составляющая половину значащих цифр числа*. Такой вывод не является для нас неожиданным (по крайней мере, качественно), поскольку потеря первых слева значащих цифр при вычитании близких значений функции при составлении таблиц конечных разностей с малым шагом уже обсуждалась в гл.1; здесь же эти разности еще делятся на малый шаг, что ведет к увеличению абсолютной погрешности приближенных значений производной.

Для формул численного дифференцирования более высоких порядков точности ситуация с потерей значащих разрядов в результатах их применения обстоит несколько лучше. Действительно, рассмотрим в тех же условиях, например, формулу симметричной аппроксимации второго порядка точности для $f'(x_i)$.

Полагая $M_3 := \max|f'''(x)|$, согласно (6.18), аналогично предыдущему имеем:

$$\left| f'(x_i) - \frac{y_{i+1} - y_{i-1}}{2h} \right| \leq g(h) := \frac{\delta}{h} + \frac{M_3}{6} h^2,$$

откуда, решая уравнение

$$-\frac{\delta}{h^2} + \frac{M_3}{3} h = 0,$$

находим $h_{opt} = \sqrt[3]{\frac{3\delta}{M_3}}$. Теперь, если посчитать для простоты

$$M_3 = 3, \quad \delta = 2^{-m}, \quad \text{получим} \quad h_{opt} = \delta^{\frac{1}{3}} = 2^{-\frac{m}{3}} \quad \text{и}$$

$$g_{min} = g(h_{opt}) = \frac{2^{-m}}{2^{-\frac{m}{3}}} + \frac{1}{2} \cdot 2^{-\frac{2}{3}m} = \frac{3}{2} \cdot 2^{-\frac{2}{3}m}.$$

Как видим, в этой ситуации можно рассчитывать на сохранение в приближенной производной примерно двух третей двоичных разрядов, имевшихся в исходных значениях функции.

Симметричная формула второго порядка точности для второй производной (6.20) в реальных условиях обеспечивает сохранение половины разрядов, как и рассмотренная выше формула первого порядка точности для первой производной. Действительно, с $\max|f''''(x)| = M_4$ из (6.20) имеем оценку

$$\left| f''(x_i) - \frac{y_{i-1} - 2y_i + y_{i+1}}{h^2} \right| \leq \frac{4\delta}{h^2} + \frac{M_4}{12} h^2 =: g(h),$$

наименьшее значение которой реализуется при $h_{opt} = 2 \cdot \sqrt[4]{\frac{3\delta}{M_4}}$.

При $M_4 = 3$, $\delta = 2^{-m}$ получаем $h_{opt} = 2 \cdot 2^{-\frac{m}{4}}$ и $g_{min} = 2 \cdot 2^{-\frac{m}{2}}$ (ср. с формулой (6.35)).

Каждое последующее (повторное) дифференцирование влечет новые потери точности. Об этом можно судить, во-первых, по росту показателя степени h с ростом порядка производной в знаменателе формул численного дифференцирования, во-вторых, по пропаданию значащих цифр в конечных разностях увеличи-

вающихся порядков, лежащих в основе этих формул. Для получения приближенных значений удовлетворительной точности при вычислении производных достаточно высоких порядков может потребоваться громадный запас знаков в значениях функции, что на практике трудно обеспечить.

Замечание 6.2. [48] Посмотрим с той же точки зрения на формулы численного интегрирования. Пусть интеграл $I \doteq \int_{-1}^1 f(x)dx$ вычисляется с помощью какой-либо квадратурной формулы $I \approx \sum_i A_i y_i$, и пусть, как и выше, $f(x_i) \approx y_i (\pm \delta)$. Тогда, так как $\sum_i A_i = 2$ (в общем случае, при интегрировании на промежутке $[a, b]$ должно быть $\sum_i A_i = b - a$), то

$$\begin{aligned} \left| I - \sum_i A_i y_i \right| &= \left| I - \sum_i A_i f(x_i) + \sum_i A_i f(x_i) - \sum_i A_i y_i \right| \leq \\ &\leq \left| I - \sum_i A_i f(x_i) \right| + \sum_i A_i |f(x_i) - y_i| \leq \left| I - \sum_i A_i f(x_i) \right| + 2\delta. \end{aligned}$$

Последнее равенство характеризует *квадратурный процесс* как *численно устойчивый*: независимо от числа используемых для приближенного вычисления интеграла значений функции, содержащих ошибки в пределах интервала $(-\delta, \delta)$, связанная с этим ошибка результата оценивается величиной $O(\delta)$, т.е. того же порядка, а это позволяет заботиться лишь об обеспечении требований к малости $\varepsilon > 0$ погрешности применяемой квадратурной формулы (разумеется, при условии $\delta < \varepsilon$).

Невозможность сделать погрешность аппроксимации производных разностными отношениями сколь угодно малой даже при убывании погрешностей вычисления значений функции вынуждает считать задачу численного дифференцирования *некорректной*. Здесь нарушается одно из условий корректности, а именно, нет непрерывной зависимости точности результата от точности входных данных в том смысле, что в процессе $h \rightarrow 0$ ошибка в значениях функции может убывать, а ошибка результата — бесконечно расти [39].

Пусть значение функции $f(x)$ в точке x представляется последовательностью приближений $y_k(x)$ с ошибками,

уменьшающимися при $k \rightarrow \infty$ по закону $\frac{\sin(k^2 x)}{k}$, т.е.*)

$$f(x) = y_k(x) + \frac{\sin(k^2 x)}{k} \quad \forall k \in \mathbb{N}. \quad (6.36)$$

Тогда, если при использовании формулы правой аппроксимации производной

$$f'(x) \approx \frac{f(x+h) - f(x)}{h}$$

вместо $f(x)$ будут подставляться приближения $y_k(x)$, т.е. будет применяться формула

$$f'(x) \approx y'_k(x) := \frac{y_k(x+h) - y_k(x)}{h},$$

то порождаемая такой подменой ошибка на основе (6.36) при каждом $k \in \mathbb{N}$ будет составлять величину

$$\begin{aligned} E &:= \frac{\sin(k^2(x+h)) - \sin(k^2 x)}{kh} = \\ &= \frac{2 \sin\left(\frac{k^2 h}{2}\right) \cos\left(k^2\left(x + \frac{h}{2}\right)\right)}{kh} = k \cdot \frac{\sin \frac{k^2 h}{2}}{\frac{k^2 h}{2}} \cdot \cos\left(k^2\left(x + \frac{h}{2}\right)\right). \end{aligned}$$

Последнее представление показывает, что если $k \rightarrow \infty$ и $h \rightarrow 0$ так, что $k^2 h \rightarrow 0$, то величина ошибки E может неограниченно возрастать. Следовательно, $y'_k(x) \nrightarrow f'(x)$.

УПРАЖНЕНИЯ

6.1. Выведите частные формулы аппроксимации второй производной в равноотстоящих узлах, основываясь на кубической интерполяции (например, из общей конечноразностной формулы (6.11)).

6.2. Запишите симметричную формулу четвертого порядка точности для аппроксимации $f'(x_i)$ и выведите ее остаточный член, используя формулу Тейлора.

* Обратим внимание, что здесь, в отличие от рассуждений, проводившихся в начале этого параграфа, мы следим за ошибками, а не за оценками их модулей.

6.3. Убедитесь, что выведенные в § 6.2 с помощью формулы Тейлора выражения остаточных членов простейших аппроксимаций первой и второй производных являются частными случаями общих формул (6.28), (6.29).

6.4. Выведите общую формулу численного дифференцирования, привлекая интерполяционную формулу Ньютона для неравных промежутков (1.43).

6.5. Непосредственной подстановкой убедитесь, что приближенная формула

$$f''(0) \approx \frac{f(-h) - 2f(0) + f(h)}{h^2}$$

точна для произвольного многочлена второй степени.

6.6. Запишите формулу полной погрешности и найдите выражение оптимального шага таблицы значений функции $f(x)$ при аппроксимации первой производной по формуле несимметричной аппроксимации второго порядка точности, считая, что абсолютные погрешности значений функции в узлах не превосходят δ и что $\max|f'''(x)| \leq M_3$.

6.7. Укажите значения оптимального шага и наивысшую гарантированную точность аппроксимации первой производной функции $y = \sin x$ посредством вычисляемых с предельной абсолютной погрешностью $\delta = 10^{-6}$ ее значений по различным простейшим формулам первого и второго порядков точности.

6.8. Найдите величины оптимального шага и оцените наилучшую точность, которая может быть гарантирована в условиях примера 6.1 (см. § 6.2) для различных рассмотренных там аппроксимаций первой и второй производных.

ГЛАВА 7 | МЕТОДЫ ЭЙЛЕРА И РУНГЕ–КУТТЫ РЕШЕНИЯ НАЧАЛЬНЫХ ЗАДАЧ ДЛЯ ОБЫКНОВЕННЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ

Ставится задача получения приближенных решений обыкновенных дифференциальных уравнений первого порядка с заданными начальными условиями. Уделяется внимание различным способам вывода численного метода Эйлера, являющегося наиболее простым частным случаем нескольких групп численных процессов разной идеологии. Рассматриваются непосредственные модификации методов Эйлера; одна из таких модификаций, называемая исправленным методом Эйлера, приводит к семейству методов Рунге–Кутты второго порядка. Записывается общий вид формул Рунге–Кутты произвольного порядка, дается геометрическая интерпретация классического метода Рунге–Кутты четвертого порядка, обсуждаются вопросы пошагового контроля точности при реализации методов, приводится алгоритм Кутты–Мерсона, решающий проблему автоматического выбора расчетного шага при численном интегрировании поставленной начальной задачи с заданной точностью.

§ 7.1. ПОСТАНОВКА ЗАДАЧИ. КЛАССИФИКАЦИЯ ПРИБЛИЖЕННЫХ МЕТОДОВ. МЕТОД ПОСЛЕДОВАТЕЛЬНЫХ ПРИБЛИЖЕНИЙ

Будем рассматривать обыкновенное дифференциальное уравнение (сокращенно ОДУ) первого порядка

$$y' = f(x, y), \quad x \in [x_0, b] \quad (7.1)$$

с начальным условием

$$y(x_0) = y_0, \quad (7.2)$$

где $f(x, y)$ — некоторая заданная, в общем случае, нелинейная функция двух переменных. Будем считать, что для данной задачи (7.1)–(7.2), называемой *начальной задачей* или *задачей Коши*, выполняются требования, обеспечивающие существование и единственность на отрезке $[x_0, b]$ ее решения $y = y(x)$ (такие требования можно найти в любом курсе дифференциальных уравнений или в соответствующем разделе курса высшей математики, см., например [59, 82, 83]). Более того, не оговаривая это отдельно, будем предполагать, что искомое решение обладает

той или иной степенью гладкости, необходимой для построения и «законного» применения того или иного метода.

Несмотря на внешнюю простоту уравнения (7.1), решить его аналитически, т.е. найти общее решение $y = y(x, C)$ с тем, чтобы затем выделить из него интегральную кривую $y = y(x)$, проходящую через заданную точку $(x_0; y_0)$, удастся лишь для некоторых специальных типов таких уравнений, описание которых также можно обнаружить, например, в упомянутых литературных источниках. Поэтому, как и в родственной для (7.1)–(7.2) задаче вычисления интегралов, приходится делать ставку на приближенные способы решения начальных задач для ОДУ, которые можно разделить на три группы:

- 1) *приближенно-аналитические методы*;
- 2) *графические или машинно-графические методы*;
- 3) *численные методы*.

К методам первой группы относят такие, которые позволяют находить приближение решения $y(x)$ сразу в виде некоторой «хорошей» функции $\varphi(x)$. Например, широко известен **метод степенных рядов**, в одну из реализаций которого заложено представление искомой функции $y(x)$ отрезком ряда Тейлора, где тейлоровские коэффициенты, содержащие производные высших порядков, находят последовательным дифференцированием самого уравнения (7.1) [31, 55 и др.]. Другим представителем этой группы методов является метод последовательных приближений, суть которого приведена чуть ниже.

Название **графические методы** говорит о приближенном представлении искомого решения $y(x)$ на промежутке $[x_0, b]$ в виде графика, который можно строить по тем или иным правилам, связанным с графическим толкованием данной задачи. Физическая или, возможно, точнее будет сказать, электротехническая интерпретация начальных задач для определенных видов уравнений лежит в основе машинно-графических методов приближенного решения. Реализуя на физико-техническом уровне заданные электрические процессы, на экране осциллографа наблюдают поведение решений дифференциальных уравнений, описывающих эти процессы. Изменение параметров уравнений приводит к адекватному изменению поведения решений, что положено в основу специализированных **аналоговых вычислительных машин** (АВМ).

Наконец, наиболее значимыми в настоящее время, характеризующее бурным развитием и проникновением во все сферы человеческой деятельности цифровой вычислительной техники, являются численные методы решения дифференциальных уравнений, предполагающие получение числовой таблицы

приближенных значений y_i искомого решения $y(x)$ на некоторой сетке $x_i \in [x_0, b]$ значений аргумента x . Этим способам и будет посвящено дальнейшее изложение. Что делать с получаемыми численными значениями решения, зависит от прикладной постановки задачи. Если речь идет о нахождении только значения $y(b)$, тогда точка b включается как конечная в систему расчетных точек x_i , и все приближенные значения $y_i \approx y(x_i)$, кроме последнего, участвуют лишь как промежуточные, т.е. не требуют запоминания, ни обработки. Если же нужно иметь приближенное решение $y(x)$ в любой точке x , то для этого к получаемой числовой таблице значений y_i можно применить какой-либо из способов аппроксимации табличных функций, рассмотренных ранее, например, интерполяцию (гл.1) или сплайн-интерполяцию (гл.4). Возможны и другие использования численных данных о решении.

Коснемся одного приближенно-аналитического способа решения начальной задачи (7.1)–(7.2), в котором искомое решение $y = y(x)$ в некоторой правой окрестности точки x_0 является пределом последовательности получаемых определенным образом функций $y_n(x)$.

Проинтегрируем левую и правую части уравнения (7.1) в границах от x_0 до x :

$$\int_{x_0}^x y'(t) dt = \int_{x_0}^x f(t, y(t)) dt.$$

Отсюда, с учетом того, что одной из первообразных для $y'(x)$ служит $y(x)$, получаем

$$y(x) - y(x_0) = \int_{x_0}^x f(t, y(t)) dt$$

или, с использованием начального условия (7.2),

$$y(x) = y_0 + \int_{x_0}^x f(t, y(t)) dt. \quad (7.3)$$

Таким образом, данное дифференциальное уравнение (7.1) с начальным условием (7.2) преобразовалось в **интегральное уравнение** (неизвестная функция здесь входит под знак интеграла; несколько более подробно об этом см. гл.11).

Полученное интегральное уравнение (7.3) имеет вид задачи о неподвижной точке

$$y = \phi(y)$$

для оператора $\phi(y) := y_0 + \int_{x_0}^x f(t, y(t)) dt$. Формально к этой задаче можно применить метод простых итераций

$$y_{n+1} = \phi(y_n), \quad n=0, 1, 2, \dots, \quad (7.4)$$

достаточно обстоятельно рассматривавшийся в предыдущей книге автора [18, гл. 3, 5, 6] применительно к скалярным уравнениям, а также к системам линейных и нелинейных алгебраических и трансцендентных уравнений. Беря в качестве начальной функции $y_0(x)$ заданную в (7.2) постоянную y_0 , по формуле (7.4) при $n = 0$ находим первое приближение

$$y_1(x) = y_0 + \int_{x_0}^x f(t, y_0) dt.$$

Его подстановка в (7.4) при $n = 1$ дает второе приближение

$$y_2(x) = y_0 + \int_{x_0}^x f(t, y_1(t)) dt,$$

и т.д. Таким образом, этот приближенно-аналитический метод, называемый *методом последовательных приближений* или *методом Пикара*^{*}, определяется формулой

$$y_{n+1}(x) = y_0 + \int_{x_0}^x f(t, y_n(t)) dt, \quad (7.5)$$

где $n = 0, 1, 2, \dots$ и $y_0(x) \equiv y_0$.

Хорошо изучена сходимость метода (7.5). В частности, не вдаваясь в подробности, констатируем один качественный результат [31]:

если в некоторой односвязной области G , содержащей точку $(x_0; y_0)$,

$$|f(x, y)| \leq C, \quad |f'_y(x, y)| \leq C_1,$$

то найдется такая постоянная $h > 0$, что на отрезке $[x_0, x_0 + h]$ последовательность функций $y_n(x)$, определяемая методом (7.5), равномерно сходится к решению $y(x)$ задачи Коши (7.1)–(7.2) и справедлива оценка погрешности

$$|y(x) - y_n(x)| \leq C \cdot C_1^n \cdot \frac{(x - x_0)^{n+1}}{(n+1)!}.$$

^{*} Пикар Шарль Эмиль (1856–1941) — французский математик.

Отметим две характеристики метода последовательных приближений Пикара, которые можно отнести к негативным. Во-первых, в силу известных проблем с эффективным нахождением первообразных, в чистом виде метод (7.5) редко реализуем. Во-вторых, как видно из вышеприведенного утверждения, этот метод следует считать локальным, пригодным для приближения решения в малой правой окрестности начальной точки. Большее значение метод Пикара имеет для доказательства существования и единственности решения задачи Коши, нежели для его практического нахождения.

§ 7.2. МЕТОД ЭЙЛЕРА — РАЗНЫЕ ПОДХОДЫ К ПОСТРОЕНИЮ

Учитывая ключевую позицию, которую занимает метод Эйлера в теории численных методов ОДУ, рассмотрим несколько способов его вывода. При этом будем считать, что вычисления проводятся с *расчетным шагом* $h = \frac{b - x_0}{n}$, *расчетными точками (узлами)* служат точки $x_i = x_0 + ih$ ($i = 0, 1, \dots, n$) промежутка $[x_0, b]$ и целью является построение таблицы

x	x_0	x_1	$\dots$	$x_n = b$
y	y_0	y_1	$\dots$	$y_n \approx y(b)$

приближенных значений y_i решения $y = y(x)$ задачи (7.1)–(7.2) в расчетных точках x_i .

Геометрический способ. Пользуясь тем, что в точке x_0 известно и значение решения $y(x_0) = y_0$ (согласно (7.2)), и значение его производной $y'(x_0) = f(x_0, y_0)$ (согласно (7.1)), можно записать уравнение касательной к графику искомой функции $y = y(x)$ в точке $(x_0; y_0)$:

$$y = y_0 + f(x_0, y_0)(x - x_0). \quad (7.6)$$

При достаточно малом шаге h ордината

$$y_1 = y_0 + hf(x_0, y_0) \quad (7.7)$$

этой касательной, полученная подстановкой в правую часть (7.6) значения $x_1 = x_0 + h$, по непрерывности должна мало отличаться от ординаты $y(x_1)$ решения $y(x)$ задачи (7.1)–(7.2). Следовательно, точка (x_1, y_1) пересечения касательной (7.6) с прямой

$x = x_1$ может быть приближенно принята за новую начальную точку. Через эту точку снова проведем прямую

$$y = y_1 + f(x_1, y_1)(x - x_1),$$

которая уже приближенно отражает поведение касательной к $y = y(x)$ в точке $(x_1; y(x_1))$. Подставляя сюда $x = x_2 (= x_1 + h)$, иначе, пересекая эту «касательную» прямой $x = x_2$, получим приближение значения $y(x_2)$ значением

$$y_2 = y_1 + hf(x_1, y_1),$$

и т.д. В итоге этого процесса, определяемого формулой

$$y_{i+1} = y_i + hf(x_i, y_i), \quad i = 0, 1, \dots, n \quad (7.8)$$

и называемого **методом Эйлера**, график решения $y = y(x)$ данной задачи Коши (7.1)–(7.2) приближенно представляется ломаной, составленной из отрезков приближенных касательных (рис. 7.1), откуда происходит другое название метода (7.8) — **метод ломаных**.

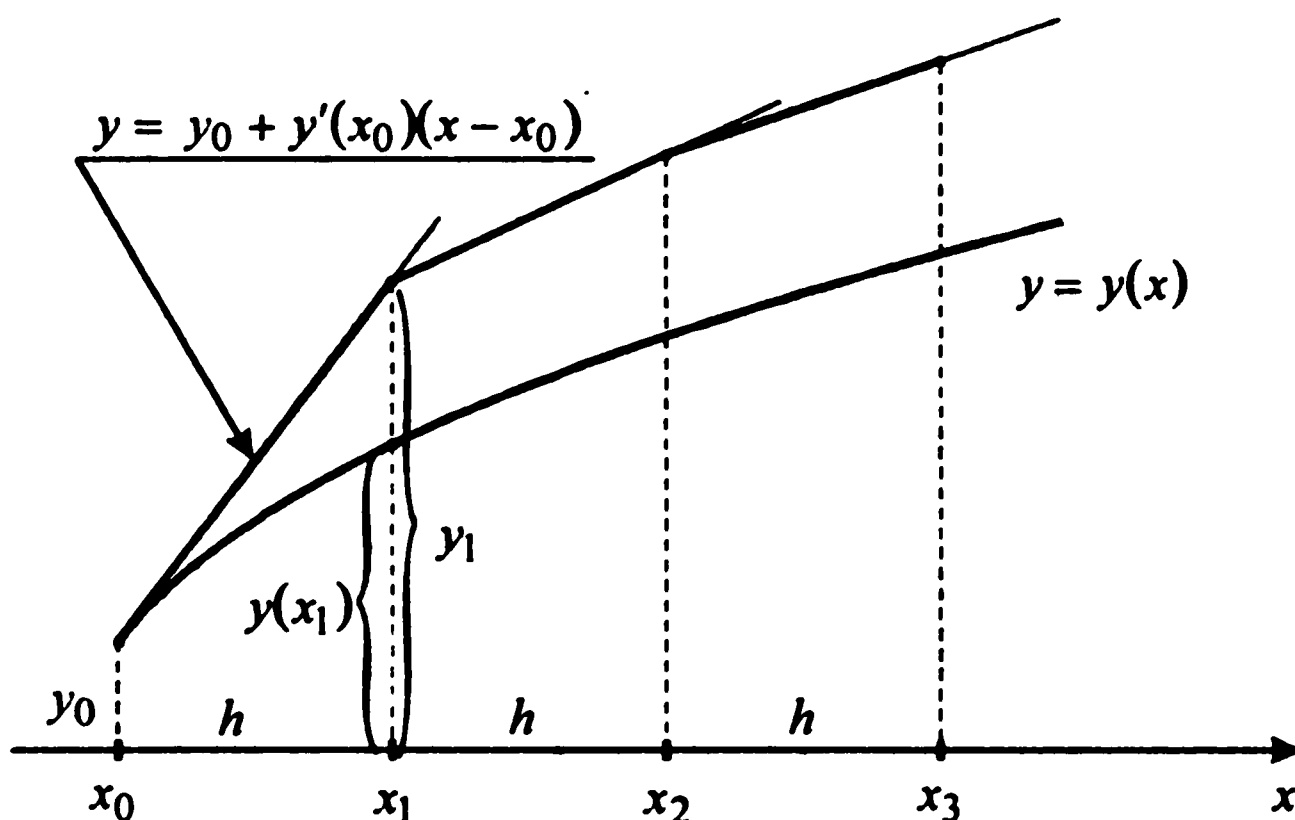


Рис. 7.1. Геометрическая интерпретация метода Эйлера

Применение формулы Тейлора. Описываемый здесь способ вывода метода Эйлера тесно связан с предыдущим. Линеаризуя решение в окрестности начальной точки по формуле Тейлора, имеем

$$y(x) = y(x_0) + y'(x_0)(x - x_0) + \frac{y''(\xi)}{2}(x - x_0)^2.$$

Отсюда при $x = x_1$ получаем

$$y(x_1) = y_0 + hf(x_0, y_0) + \frac{y''(\xi_1)}{2} h^2. \quad (7.9)$$

Точное равенство (7.9), переписанное в виде

$$y(x_1) = y_1 + r_1(h),$$

говорит о том, что здесь мы имеем одновременно как саму формулу Эйлера для вычисления значения $y_1 \approx y(x_1)$ (ср. с формулой (7.7)), так и ее остаточный член

$$r_1(h) = \frac{y''(\xi_1)}{2} h^2, \quad (7.10)$$

где ξ_1 — некоторая точка интервала (x_0, x_1) .

Остаточный член (7.10) характеризует *локальную* (или, иначе, *шаговую*) *ошибку* метода Эйлера, т.е. ошибку, совершаемую на одном шаге. Очевидно, что от шага к шагу, т.е. при многократном применении формулы (7.8), возможно наложение ошибок^{*)}. За n шагов, т.е. в точке b , образуется *глобальная ошибка*; изучение такой ошибки будет проведено позже (см. § 9.2). Сейчас же анонсируем один важный хотя бы для терминологии факт: *порядок глобальной ошибки (относительно шага h) на единицу ниже, чем порядок локальной ошибки, а порядком глобальной ошибки и определяется порядок соответствующего численного процесса решения задачи Коши.* Таким образом, локальная ошибка метода Эйлера, согласно (7.10), есть $O(h^2)$, глобальная — $O(h)$, т.е. *метод Эйлера относится к методам первого порядка.*

Разностный способ. Рассматривая уравнение (7.1) в точке x_0 , с учетом (7.2) имеем равенство

$$y'(x_0) = f(x_0, y_0).$$

Применяя к его левой части аппроксимацию производной правым разностным отношением первого порядка точности

^{*)} Заметим, что в отличие от метода последовательных приближений (7.5), являющегося итерационным по сути, все рассматриваемые численные процессы решения задачи Коши для ОДУ, и в частности, метод Эйлера (7.8), являются итерационными лишь по форме. На самом деле, это *шаговые* методы, в которых на каждом шаге выполняются однотипные действия; характерного для итерационных методов уточнения решения здесь не происходит.

(см. (6.15) при $i = 0$)

$$y'(x_0) = \frac{y(x_1) - y(x_0)}{h} - \frac{y''(\xi_1)}{2}h,$$

получаем

$$\frac{y(x_1) - y(x_0)}{h} = f(x_0, y_0) + \frac{y''(\xi_1)}{2}h,$$

что идентично равенству (7.9), поставляющему формулу для вычисления y_1 вида (7.7) и локальный остаточный член (7.10). Ясно, что для получения общей расчетной формулы (7.8) можно было сразу применить аппроксимацию $y'(x_i)$ по формуле (6.16) в равенстве

$$y'(x_i) = f(x_i, y(x_i)), \quad (7.11)$$

заменив неизвестное точное значение $y(x_i)$ известным приближенным значением y_i .

Заметим, что *порядок получающегося таким способом метода численного интегрирования дифференциальной задачи (7.1)–(7.2) совпадает с порядком аппроксимации производной в левой части уравнения (7.1).*

Квадратурный способ. Как было показано в § 7.1, начальную задачу для ОДУ (7.1)–(7.2) можно заменить эквивалентным интегральным уравнением (7.3). При $x = x_1$ из него получится равенство

$$y(x_1) = y_0 + \int_{x_0}^{x_1} f(x, y(x))dx. \quad (7.12)$$

Применение к интегралу в правой части равенства (7.12) простейшей (одноточечной) формулы левых прямоугольников (5.6) дает приближенную формулу

$$y(x_1) \approx y_0 + f(x_0, y(x_0))(x_1 - x_0),$$

правая часть которой, очевидно, совпадает с выражением (7.7) для подсчета значения y_1 . В общем случае расчетная формула (7.8) метода Эйлера получается численным интегрированием посредством простейшей формулы левых прямоугольников в равенстве

$$y(x_{i+1}) - y(x_i) = \int_{x_i}^{x_{i+1}} f(x, y(x))dx \quad (7.13)$$

В предположении, что на каждом i -м шаге в роли начальной точки (x_0, y_0) выступает точка (x_i, y_i) . Зная точность исполь-

зуемой в (7.13) квадратурной формулы, легко прийти к тому же выражению локальной ошибки метода Эйлера, что и при других способах его построения.

Существуют и другие подходы к выводу метода Эйлера. В частности, он будет возникать далее как частный случай некоторых семейств численных методов решения задачи (7.1)–(7.2).

§ 7.3. НЕСКОЛЬКО ПРОСТЫХ МОДИФИКАЦИЙ МЕТОДА ЭЙЛЕРА

Разовьем последний из подходов к построению метода Эйлера. Очевидно, применение к интегральному равенству (7.13) других простейших квадратурных формул будет порождать новые методы численного интегрирования задачи Коши (7.1)–(7.2).

Так, если в (7.13) использовать простейшую квадратурную формулу правых прямоугольников (5.7), придем к методу

$$y_{i+1} = y_i + hf(x_{i+1}, y_{i+1}), \quad i = 0, 1, \dots, n. \quad (7.14)$$

Этот метод называют *неявным* (или *обратным*) *методом Эйлера*, поскольку для вычисления неизвестного значения $y_{i+1} \approx y(x_{i+1})$ по известному значению $y_i \approx y(x_i)$ требуется решать уравнение, в общем случае нелинейное. Имеет ли свою сферу применения подобный метод, порядок которого такой же, как и у явного метода Эйлера (7.8) (первый^{*)}), и один шаг вычислений по которому столь трудоемок? Положительный ответ на этот вопрос будет дан в гл. 9.

Применение к интегралу в (7.13) простейшей квадратурной формулы трапеций (5.26) приводит тоже к неявному методу

$$y_{i+1} = y_i + \frac{h}{2}[f(x_i, y_i) + f(x_{i+1}, y_{i+1})], \quad i = 0, 1, \dots, n, \quad (7.15)$$

который будем называть *методом трапеций*. Квадратурная формула трапеций, как известно из гл.5, на порядок точнее формул левых и правых прямоугольников. Отсюда вытекает более высокий (на единицу) порядок точности метода трапеций (7.15) по сравнению с явным и с неявным методами Эйлера (7.8) и (7.14), т.е. *метод трапеций* (7.15) — это метод второго порядка. (Впоследствии, в § 8.4, к этому выводу придем из других соображений.)

Некоторый интерес представляет совместное применение явного метода Эйлера и неявного метода трапеций.

^{*)} Очевидно, локальный остаточный член метода (7.14) лишь знаком будет отличаться от локального остаточного члена (7.10) метода (7.8).

По форме равенство (7.15) представляет собой скалярную задачу о неподвижной точке относительно неизвестного значения y_{i+1} . Поэтому, если в правую часть (7.15) подставить хорошее начальное приближение y_{i+1}^0 , подсчитываемое по формуле (7.14), то тогда само это равенство можно считать шагом метода простых итераций для уточнения этого значения [18, гл.5]. Таким образом, получаем гибридный метод

$$y_{i+1} = y_i + \frac{h}{2} [f(x_i, y_i) + f(x_i + h, y_i + hf(x_i, y_i))], \quad (7.16)$$

$$i = 0, 1, \dots, n,$$

который называют *методом Хойна* [4, 29] (или *Хьюна* [73]).

Ясно, что можно достичь большей точности, если, исходя из того же начального приближения

$$y_{i+1}^0 = y_i + hf(x_i, y_i),$$

сделать не одну, а несколько итераций по методу трапеций:

$$y_{i+1}^{(k)} = y_i + \frac{h}{2} [f(x_i, y_i) + f(x_{i+1}, y_{i+1}^{(k-1)})], \quad k = 1, 2, \dots \quad (7.17)$$

Такой вариант совместного применения метода Эйлера и метода трапеций называют *усовершенствованным методом Эйлера-Коши с итерационной обработкой*^{*}. Делать много итераций по формуле (7.17) не рекомендуется (обычно их выполняют не более трех-четырех). Совпадение определенного числа разрядов в полученных числах $y_{i+1}^{(k)}$ и $y_{i+1}^{(k-1)}$ говорит о точности, с которой решено методом простых итераций уравнение (7.15) относительно y_{i+1} , а вовсе не о том, что с такой точностью найдено значение $y(x_{i+1})$.

Чтобы получить следующую модификацию метода Эйлера, проинтегрируем уравнение (7.1) по отрезку $[x_{i-1}, x_{i+1}]$. Имеем

$$\int_{x_{i-1}}^{x_{i+1}} y'(x) dx = \int_{x_{i-1}}^{x_{i+1}} f(x, y(x)) dx,$$

откуда следует равенство

$$y(x_{i+1}) = y(x_{i-1}) + \int_{x_{i-1}}^{x_{i+1}} f(x, y(x)) dx. \quad (7.18)$$

Применяя к последнему интегралу одноточечную квадратурную формулу средних прямоугольников (5.10) и заменяя значения $y(x_{i-1})$ и $y(x_i)$ известными приближенными значениями y_{i-1} и

^{*} Иначе, *методом Эйлера с пересчетом* [6].

y_i соответственно, из (7.18) выводим формулу для подсчета приближенного значения $y(x_{i+1})$

$$y_{i+1} = y_{i-1} + 2hf(x_i, y_i), \quad i = 1, 2, \dots, n-1, \quad (7.19)$$

которую будем называть **уточненным методом Эйлера** [27].

Как известно, квадратурная формула прямоугольников (средней точки) имеет тот же порядок точности, что и квадратурная формула трапеций, так что **уточненный метод Эйлера** (7.19) тоже является **методом второго порядка**^{*)}. Подтверждением этого факта может служить вывод метода (7.19) на разностной основе. Применив к равенству (7.11) формулу симметричной аппроксимации $y'(x_i)$ второго порядка точности, получим

$$\frac{y(x_{i+1}) - y(x_{i-1}))}{2h} \approx f(x_i, y(x_i)),$$

откуда после приближенной замены $y(x_{i-1}) \approx y_{i-1}$, $y(x_i) \approx y_i$, $y(x_{i+1}) \approx y_{i+1}$ следует (7.19).

Обратим внимание на одно принципиальное отличие метода (7.19) от всех других рассмотренных до этого момента методов: метод (7.19) является **двухшаговым**. Здесь для вычисления значения y_{i+1} привлекаются два предыдущих значения y_i и y_{i-1} . Двухшаговость накладывает определенные ограничения, по крайней мере, на начало численного процесса: значение $y_1 \approx y(x_1)$ не может быть найдено непосредственно этим методом с тем же шагом h . Поэтому недостающую вторую начальную для процесса (7.19) точку приходится получать другим путем, например, явным методом Эйлера, а чтобы не сделать сразу большой ошибки, применяя на старте метод более низкого порядка точности, рекомендуется осуществлять постепенное вхождение в процесс (7.19). Так, «разгон» можно выполнить по формулам

$$y_{\frac{1}{2}} = y_0 + \frac{h}{2}f(x_0, y_0), \quad y_1 = y_0 + hf\left(x_0 + \frac{h}{2}, y_{\frac{1}{2}}\right), \quad (7.20)$$

а далее уже переключаться на счет по формуле (7.19).

Пример 7.1. Рассмотрим простое линейное уравнение

$$y' = 2x - 3y$$

с начальным условием $y(0) = 1$. На этой задаче легко проследить за вы-

^{*)} Другие его названия: **метод Нюстрёма** (Нюстрема) **второго порядка** [4, 89], **метод Милна второго порядка** [95].

числениями, реализующими различные выведенные выше методы. Знание ее точного решения $y(x) = -\frac{2}{9} + \frac{2}{3}x + \frac{11}{9}e^{-3x}$ позволяет провести сравнение результатов приближенных вычислений по разным формулам с истинным решением и проверить, насколько соответствуют представления о точности тех или иных методов тому, что наблюдается в данном, наверное, далеко не самом типичном частном случае.

Сначала сделаем несколько последовательных приближений по методу Пикара. Его итерационная формула (7.5) для данной начальной задачи имеет вид

$$y_{n+1}(x) = 1 + x^2 - 3 \int_0^x y_n(t) dt.$$

Подставляя сюда $y_0 = 1$, при $n = 0, 1, 2$ последовательно получаем:

$$y_1(x) = 1 + x^2 - 3 \int_0^x dt = 1 - 3x + x^2,$$

$$y_2(x) = 1 + x^2 - 3 \int_0^x (1 - 3t + t^2) dt = 1 - 3x + \frac{11}{2}x^2 - x^3,$$

$$y_3(x) = 1 + x^2 - 3 \int_0^x \left(1 - 3t + \frac{11}{2}t^2 - t^3\right) dt = 1 - 3x + \frac{11}{2}x^2 - \frac{11}{2}x^3 + \frac{3}{4}x^4.$$

Эти результаты удобно сравнить с точным решением, если в последнем разложить в ряд по степеням x фигурирующую там функцию e^{-3x} . Тогда получим представление решения в виде ряда

$$y(x) = 1 - 3x + \frac{11}{2}x^2 - \frac{11}{2}x^3 + \frac{33}{8}x^4 - \dots,$$

с которым, как видим, хорошо согласуются приближения y_1, y_2, y_3 , определяемые методом Пикара.

Теперь проведем подсчет приближенных значений решения $y(x)$ данной задачи в точке $x = 0.2$ численным методом Эйлера и его модификациями, принимая $h = 0.1$ (т.е. за два шага). Результаты этих вычислений и фактические ошибки, найденные сравнением с точным значением $y(0.2) = 0.581881\dots$, отражены в следующей таблице.

Метод	$y_1 \approx y(0.1)$	$y_2 \approx y(0.2)$	$y(0.2) - y_2$
Метод Эйлера (7.8)	0.7	0.51	≈ 0.07
Неявный метод Эйлера (7.14)	≈ 0.7846	≈ 0.6343	≈ -0.05
Метод трапеций (7.15)	≈ 0.7478	≈ 0.5788	≈ 0.003
Метод Хойна (7.16)	0.755	≈ 0.5895	≈ -0.008
Уточненный метод Эйлера (7.19)–(7.20)	0.755	0.587	≈ -0.005

Последний столбец в этой таблице со всей очевидностью показывает большую точность методов второго порядка (см. три последних строки).

§ 7.4. ИСПРАВЛЕННЫЙ МЕТОД ЭЙЛЕРА

Пусть найдено приближенное значение $y_i \approx y(x_i)$ решения $y = y(x)$ задачи (7.1)–(7.2) и требуется вычислить $y_{i+1} \approx y(x_{i+1})$, где $x_{i+1} = x_i + h$. Запишем разложение решения по формуле Тейлора p -го порядка, принимая за базовую точку x_i (т.е. по степеням $x - x_i$) и положим в этом разложении $x = x_{i+1}$. Имеем

$$y(x_{i+1}) = y(x_i) + hy'(x_i) + \frac{1}{2!}h^2 y''(x_i) + \dots \\ \dots + \frac{1}{p!}h^p y^{(p)}(x_i) + O(h^{p+1}). \quad (7.21)$$

Если ограничиться двумя слагаемыми в правой части разложения (7.21), то, согласно показанному в § 7.2, получим обычный метод Эйлера (7.8). Посмотрим, что дает учитывание третьего слагаемого.

При $p = 2$ из (7.21) следует равенство

$$y(x_{i+1}) = y(x_i) + hy'(x_i) + \frac{h^2}{2} y''(x_i) + O(h^3). \quad (7.22)$$

Значение первой производной в точке x_i , в силу связи (7.1), приближенно известно:

$$y'(x_i) = f(x_i, y(x_i)) \approx f(x_i, y_i). \quad (7.23)$$

Дифференцируя (7.1), по формуле полной производной

$$y''(x) = f'_x(x, y) + f'_y(x, y)y'$$

находим приближенное значение второй производной:

$$y''(x_i) = f'_x(x_i, y(x_i)) + f'_y(x_i, y(x_i))f(x_i, y(x_i)) \approx \\ \approx f'_x(x_i, y_i) + f'_y(x_i, y_i)f(x_i, y_i). \quad (7.24)$$

Подставляя приближенные выражения $y(x_i)$, $y'(x_i)$ и $y''(x_i)$ в равенство (7.22), получаем следующую формулу для вычисления $y_{i+1} \approx y(x_{i+1})$ при $i = 0, 1, \dots, n$:

$$y_{i+1} = y_i + h \left[f(x_i, y_i) + \frac{h}{2} (f'_x(x_i, y_i) + f'_y(x_i, y_i)f(x_i, y_i)) \right]. \quad (7.25)$$

Определяемый ею метод будем называть *исправленным методом Эйлера*.

Так как при $i = 0$ формулы (7.23) и (7.24) точны, а $y_0 = y(x_0)$, согласно начальному условию (7.2), то на первом шаге вычислений по формуле (7.25) будет совершаться ошибка, связанная только с усечением ряда Тейлора. Следовательно, локальная ошибка или, иначе, **шаговая погрешность** метода (7.25) составляет величину $O(h^3)$, а это означает, что *исправленный метод Эйлера относится к методам второго порядка*.

§ 7.5. О СЕМЕЙСТВЕ МЕТОДОВ РУНГЕ–КУТТЫ. МЕТОДЫ ВТОРОГО ПОРЯДКА

Недостатком исправленного метода Эйлера (7.25) и других методов более высоких порядков, основанных на пошаговом представлении решения $y(x)$ задачи (7.1)–(7.2) по формуле Тейлора и последовательном дифференцировании уравнения (7.1) для получения тейлоровых коэффициентов, является необходимость вычисления на каждом шаге частных производных функции $f(x, y)$.

Идея построения явных **методов Рунге–Кутты**^{*)} p -го порядка заключается в получении приближений к значениям $f(x_{i+1})$ по формуле вида

$$y_{i+1} = y_i + h\varphi(x_i, y_i, h), \quad (7.26)$$

где $\varphi(x, y, h)$ — некоторая функция, приближающая отрезок ряда Тейлора (7.21) до p -го порядка и не содержащая частных производных функции $f(x, y)$.

Так, полагая в (7.26) $\varphi(x, y, h) \equiv f(x, y)$, приходим к методу Эйлера (7.8), т.е. метод Эйлера можно считать простейшим примером методов Рунге–Кутты, соответствующим случаю $p = 1$.

Для построения методов Рунге–Кутты порядка, выше первого, функцию $\varphi(x, y, h)$ берут многопараметрической, и подбирают ее параметры сравнением выражения (7.26) с многочленом Тейлора для $y(x)$ соответствующей желаемому порядку степени.

Рассмотрим случай $p = 2$. Возьмем функцию φ в (7.26)

^{*)} 1) Кутта Мартин Вильгельм (1867–1944) — немецкий физик и математик.

2) Ранее типичным было написание «метод Рунге–Кутта»; в последнее время все чаще говорят и пишут «метод Рунге–Кутты».

3) О неявных методах Рунге–Кутты можно прочитать, например, в [91].

следующей структуры:

$$\varphi(x, y, h) := c_1 f(x, y) + c_2 f(x + ah, y + bhf(x, y)).$$

Ее параметры c_1 , c_2 , a и b будем подбирать так, чтобы записанная, согласно (7.26), формула

$$y_{i+1} = y_i + h[c_1 f(x_i, y_i) + c_2 f(x_i + ah, y_i + bhf(x_i, y_i))] \quad (7.27)$$

определяла метод второго порядка, т.е. чтобы максимальная локальная ошибка составляла величину $O(h^3)$.

Разложим функцию двух переменных $f(x + ah, y + bhf(x, y))$ по формуле Тейлора, ограничиваясь линейными членами:

$$f(x + ah, y + bhf(x, y)) = f(x, y) + f'_x(x, y)ah + \\ + f'_y(x, y)bhf(x, y) + O(h^2).$$

Ее подстановка в (7.27) дает

$$y_{i+1} = y_i + h[(c_1 + c_2)f(x_i, y_i) + h(c_2af'_x(x_i, y_i) + \\ + c_2bf'_y(x_i, y_i)f(x_i, y_i))] + O(h^3). \quad (7.28)$$

Сравнение последнего выражения с тейлоровским квадратичным представлением решения $y(x)$ (7.22) с точностью до $O(h^3)$ равносильно сравнению его с выражением y_{i+1} по формуле (7.25), т.е. с исправленным методом Эйлера. Очевидно, чтобы (7.28) и (7.25) совпадали с точностью $O(h^3)$, от параметров нужно потребовать выполнение следующей совокупности условий:

$$\begin{cases} c_1 + c_2 = 1, \\ c_2a = 0.5, \\ c_2b = 0.5. \end{cases} \quad (7.29)$$

Полученная система условий содержит три уравнения относительно четырех параметров метода. Это говорит о наличии одного свободного параметра. Положим $c_2 = \alpha (\neq 0)$. Тогда из (7.29) имеем:

$$c_1 = 1 - \alpha, \quad a = \frac{1}{2\alpha}, \quad b = \frac{1}{2\alpha}.$$

В результате подстановки этих значений параметров в формулу (7.27) приходим к *однопараметрическому семейству методов Рунге–Кутты второго порядка*.

$$y_{i+1} = y_i + h\left[(1 - \alpha)f(x_i, y_i) + \alpha f\left(x_i + \frac{h}{2\alpha}, y_i + \frac{h}{2\alpha}f(x_i, y_i)\right)\right]. \quad (7.30)$$

Выделим из семейства методов (7.30) два наиболее простых и естественных частных случая:

при $\alpha = \frac{1}{2}$ получаем формулу

$$y_{i+1} = y_i + \frac{h}{2} [f(x_i, y_i) + f(x_i + h, y_i + hf(x_i, y_i))],$$

в котором узнаём **метод Хойна** (7.16), полученный ранее из других соображений;

при $\alpha = 1$ из (7.30) выводим новый простой метод

$$y_{i+1} = y_i + hf\left(x_i + \frac{h}{2}, y_i + \frac{h}{2} f(x_i, y_i)\right), \quad (7.31)$$

который назовем **методом средней точки**.

§ 7.6. МЕТОДЫ РУНГЕ–КУТТЫ ПРОИЗВОЛЬНОГО И ЧЕТВЕРТОГО ПОРЯДКОВ

Любой метод из семейства методов Рунге–Кутты второго порядка (7.30) реализуют по следующей схеме. На каждом шаге, т.е. при каждом $i = 0, 1, 2, \dots$, вычисляют значения функции

$$\eta_1^i := f(x_i, y_i),$$

$$\eta_2^i := f\left(x_i + \frac{h}{2\alpha}, y_i + \frac{h}{2\alpha} \eta_1^i\right),$$

а затем находят **шаговую поправку**

$$\Delta y_i := h[(1 - \alpha)\eta_1^i + \alpha\eta_2^i],$$

прибавление которой к результату предыдущего шага дает приближенное значение решения $y(x)$ в точке $x_{i+1} = x_i + h$:

$$y_{i+1} = y_i + \Delta y_i.$$

Метод такой структуры называют **двухэтапным** по количеству вычислений значений функции — правой части уравнения (7.1) — на одном шаге.

Анализ устройства методов Рунге–Кутты второго порядка позволяет представить, в какой форме следует конструировать явный метод Рунге–Кутты произвольного порядка. По аналогии с предыдущим для семейства методов Рунге–Кутты p -го порядка

используется запись, состоящая из следующей совокупности формул:

$$\begin{cases} \eta_1^i = f(x_i, y_i), \\ \eta_k^i = f\left(x_i + a_k h, y_i + h \sum_{j=1}^{k-1} b_{kj} \eta_j^i\right), \\ y_{i+1} = y_i + h \sum_{k=1}^p c_k \eta_k^i, \end{cases} \quad (7.32)$$

где $k = 2, 3, \dots, p$ (для p -этапного метода). Многочисленные параметры c_k, a_k, b_{kj} , фигурирующие в формулах (7.32), подбираются так, чтобы получаемое методом (7.32) значение y_{i+1} совпадало со значением разложения $y(x_{i+1})$ по формуле Тейлора с погрешностью $O(h^{p+1})$ (без учета погрешностей, совершаемых на предыдущих шагах).

Наиболее употребительным частным случаем семейства методов (7.32) является следующий **метод Рунге–Кутты четвертого порядка**, относящийся к **четырёхэтапным** и имеющий вид^{*}):

$$\begin{cases} \eta_1^i = f(x_i, y_i), \\ \eta_2^i = f\left(x_i + \frac{h}{2}, y_i + \frac{h}{2} \eta_1^i\right), \\ \eta_3^i = f\left(x_i + \frac{h}{2}, y_i + \frac{h}{2} \eta_2^i\right), \\ \eta_4^i = f(x_i + h, y_i + h \eta_3^i), \\ \Delta y_i = \frac{h}{6} (\eta_1^i + 2\eta_2^i + 2\eta_3^i + \eta_4^i), \\ y_{i+1} = y_i + \Delta y_i. \end{cases} \quad (7.33)$$

Не пытаясь воспроизвести выкладки, приводящие от общей записи семейства (7.32) при $p = 4$ к конкретному методу (7.33), дадим геометрическое толкование последнего.

^{*} Как правило, когда говорят «метод Рунге–Кутты», не сообщая о нем никаких дополнительных сведений, то под этим подразумевают именно данный классический метод четвертого порядка.

Обратив внимание на то, что шаговая поправка Δy_i есть средневзвешенная величина поправок $h\eta_1^i, h\eta_2^i, h\eta_3^i, h\eta_4^i$ каждого этапа (с весовыми коэффициентами $\frac{1}{6}, \frac{2}{6}, \frac{2}{6}, \frac{1}{6}$ соответственно), проанализируем, как получаются эти поправки этапов. На первом этапе создается приращение $h\eta_1^i = hf(x_i, y_i)$ ($= hy'(x_i)$), соответствующее шаговой поправке Эйлера, — это очевидно. На рис 7.2 ему отвечает отрезок BC вертикали $x = x_{i+1}$ (точка B получена ортогональным проектированием точки A на эту вертикаль).

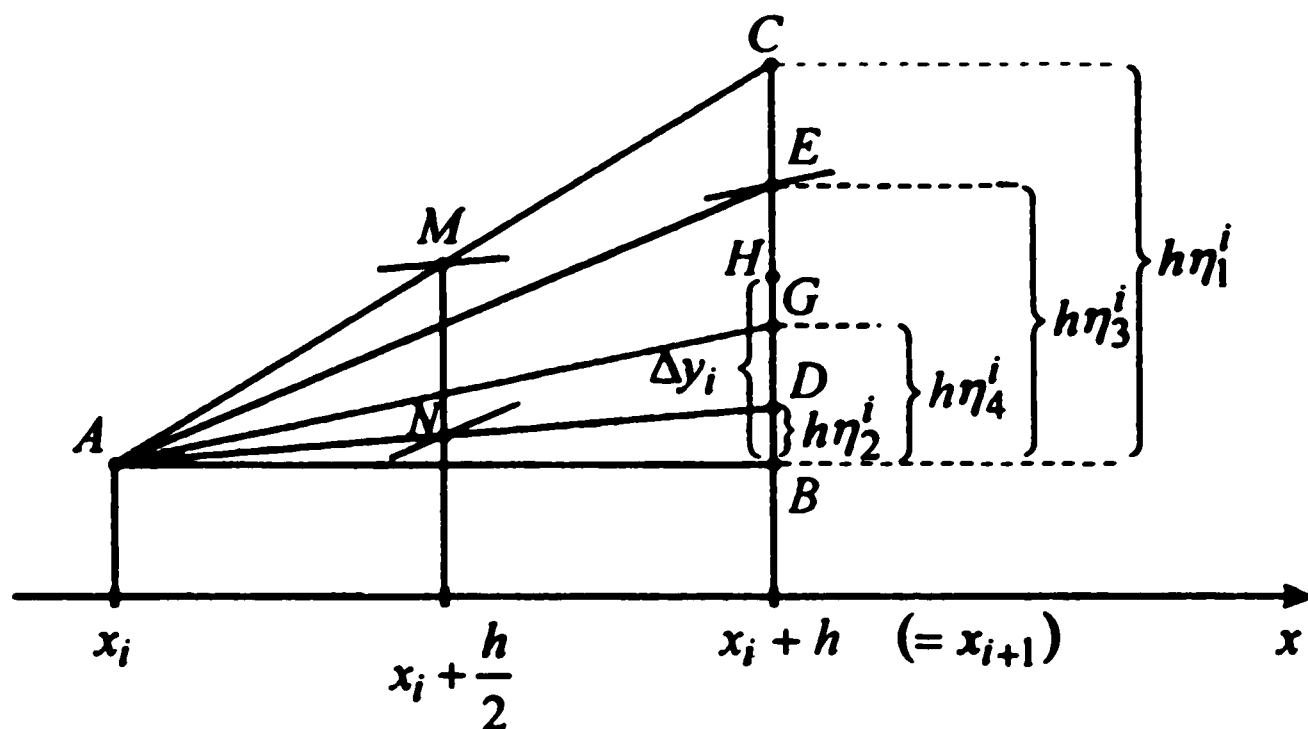


Рис. 7.2. Геометрическая иллюстрация одного шага метода Рунге–Кутты четвертого порядка

Так как точка M , благодаря свойству средней линии треугольника (см. $\triangle ABC$), имеет ординату $y_i + \frac{h}{2}\eta_1^i$, то η_2^i определяет значение $f(M)$, служащее (согласно связи $y' = f(x, y)$ и геометрическому смыслу производной) тангенсом угла A в новом треугольнике с противолежащим этому углу катетом $h\eta_2^i = BD$. Далее, аналогично, подсчитав $f(N) = f\left(x_i + \frac{h}{2}, y_i + \frac{h}{2}\eta_2^i\right) = \eta_3^i$, на вертикали $x = x_{i+1}$ откладываем следующую промежуточную (этапную) поправку $h\eta_3^i = BE$. Вычислив величину $f(E) = f(x_i + h, y_i + h\eta_3^i)$, являющуюся значением тангенса

угла A во вновь получаемом $\triangle ABG$, имеем поправку $h\eta_4^i = BG$ последнего этапа. Итоговая шаговая поправка $\Delta y_i = BH$ есть продукт усреднения с указанными коэффициентами четырех этапных поправок — длин отрезков BC , BD , BE и BG . Точка H будет стартовой для следующего, $i + 1$ -го, шага метода (7.33).

Заметим, что если первый этап, как уже упоминалось, соответствует применению явного метода Эйлера, то четвертый — неявного, а второй и третий — уточненного методов Эйлера. Последний имеет более высокий порядок точности, отсюда и больший вес отвечающих ему значений этапных поправок.

§ 7.7. ПОШАГОВЫЙ КОНТРОЛЬ ТОЧНОСТИ. МЕТОД КУТТЫ–МЕРСОНА

Нетрудно понять, что выведение надежных и, в то же время, простых и эффективных оценок погрешности, гарантирующих получение таблицы значений решения $y = y(x)$ заданной точности, является делом малоперспективным, особенно для методов более-менее высоких порядков. Поэтому главным способом отслеживания точности при реализации численных процессов решения задачи Коши остается применение различных полуэмпирических правил, основанных на принципе Рунге (см. § 5.5).

Будем считать, что при использовании метода p -го порядка абсолютная шаговая погрешность должна находиться в пределах $\varepsilon > 0$. Тогда, согласно *принципу Рунге*, осуществляется счет по системе узлов $x_i(h) = x_0 + ih$ с шагом h и по системе узлов

$x_j\left(\frac{h}{2}\right) = x_0 + j\frac{h}{2}$ с шагом $\frac{h}{2}$. При четных j вторая система будет

совпадать с первой, т.е. $x_i(h) = x_{2i}\left(\frac{h}{2}\right)$. Переход от расчетной

точки x_i с приближенным значением решения в ней y_i к расчетной точке x_{i+1} один раз совершается за один шаг длины h и приводит к значению $y_{i+1}(h) \approx y(x_{i+1}(h))$, другой раз совершается за

два шага длины $\frac{h}{2}$ («транзитом» через точку $x_i + \frac{h}{2} = x_{2i+1}\left(\frac{h}{2}\right)$ со

значением $y_{2i+1}\left(\frac{h}{2}\right) \approx y\left(x_i + \frac{h}{2}\right)$ и дает значение

$$y_{2i+2}\left(\frac{h}{2}\right) \approx y\left(x_{2i+2}\left(\frac{h}{2}\right)\right) = y(x_{i+1}(h)).$$

Поправка Ричардсона в таком случае будет составлять величину

$$R_i\left(\frac{h}{2}\right) := \frac{y_{2i+2}\left(\frac{h}{2}\right) - y_{i+1}(h)}{2^p - 1}. \quad (7.34)$$

Если величина $\left|R_i\left(\frac{h}{2}\right)\right|$ меньше заданного ε , то можно считать, что ошибка приближенного равенства $y(x_{i+1}) \approx y_{2i+2}\left(\frac{h}{2}\right)$ не превосходит ε . Если же $\left|R_i\left(\frac{h}{2}\right)\right| > \varepsilon$, то следует уменьшить расчетный шаг h . При условии $\left|R_i\left(\frac{h}{2}\right)\right| \ll \varepsilon$ стоит попытаться двигаться дальше с более крупным шагом (например, удвоить h).

Пример 7.2 (продолжение примера 7.1, см. § 7.3).

Посмотрим, что дает применение принципа Рунге к нескольким простым методам численного решения того же уравнения $y' = 2x - 3y$ с начальным условием $y(0) = 1$. Из точки $x = 0$ перейдем в точку $x = 0.2$ за один шаг $h = 0.2$ четырьмя одношаговыми методами: явным и неявным методами Эйлера, методом трапеций и методом Хойна — частным случаем метода Рунге–Кутты второго порядка. С помощью полученных значений $y_1^{h=0.2} \approx y(0.2)$ и найденных ранее теми же методами в примере 7.1 значений $y_2^{h=0.1} \approx y(0.2)$ подсчитаем поправки Ричардсона

$$R_1^{h=0.1} = \frac{y_2^{h=0.1} - y_1^{h=0.2}}{2^p - 1}$$

при $p = 1$ для методов Эйлера и $p = 2$ для методов трапеций и Хойна. Эти результаты, а также уточненные прибавлением к значениям $y_2^{h=0.1}$ поправок Ричардсона $R_1^{h=0.1}$ приближенные значения $\tilde{y}_2^{h=0.1}$ решения $y(0.2)$ и

их истинные погрешности $y(0.2) - \tilde{y}_2^{h=0.1}$ сведем в следующую таблицу.

Метод	$y_1^{h=0.2}$	$R_1^{h=0.1}$	$\tilde{y}_2^{h=0.1}$	$y(0.2) - \tilde{y}_2^{h=0.1}$	$y(0.2) - y_2^{h=0.1}$
Метод Эйлера (7.8)	0.4	0.11	0.62	≈ -0.04	≈ 0.07
Неявный метод Эйлера (7.14)	0.675	≈ -0.0407	≈ 0.5936	≈ -0.01	≈ -0.05
Метод трапеций (7.15)	≈ 0.5692	≈ 0.0032	≈ 0.5820	≈ -0.0001	≈ 0.003
Метод Хойна (7.16)	0.62	≈ -0.0102	≈ 0.5793	≈ 0.0026	≈ -0.008

В эту таблицу последним столбцом помещен последний столбец из таблицы результатов примера 7.1, содержащий погрешности значений $y_2^{h=0.1}$. Сравнение с ним столбца со значениями поправок Ричардсона показывает, что эти поправки хорошо отражают поведение погрешностей методов (хотя и не дают основания считать их модули оценками погрешностей), а предпоследнего — эффективность уточнения по правилу Рунге–Ричардсона.

Грубо обозначенная здесь технология пошагового контроля точности численного интегрирования дифференциальных уравнений и автоматического выбора расчетного шага при этом на основе двойного счета в такой непосредственной форме говорит о ее значительной «дороговизне». Действительно, предположим, что для решения задачи (7.1)–(7.2) применяется четырехэтапный метод Рунге–Кутты четвертого порядка (7.33). Тогда выполнение одного его шага с контролем точности по правилу Рунге потребует 11 вычислений правой части уравнения (7.1) (по четыре для получения каждого из значений $y_{i+1}(h)$, $y_{2i+1}(\frac{h}{2})$ и $y_{2i+2}(\frac{h}{2})$ минус одно общее для $y_{i+1}(h)$ и $y_{2i+1}(\frac{h}{2})$), что весьма затратно.

Более «дешевый», но, возможно, менее строгий способ судить о том, достаточно ли малым выбран шаг h расчетов по методу Рунге–Кутты четвертого порядка (7.33), — это вычисление при каждом $i = 0, 1, 2, \dots$ величин

$$\Theta_i = \left| \frac{\eta_2^i - \eta_3^i}{\eta_1^i - \eta_2^i} \right|.$$

Считается, что если величина Θ_i не превосходит нескольких сотых, то можно продолжить вычисления с данным шагом или пытаться при переходе от i к $i+1$ его увеличить; в противном случае шаг следует уменьшить, например, вдвое [31, 45, 91 и др.].

Борьба за вычислительную эффективность привела к появлению различных вычислительных версий методов Рунге–Кутты, благо для этого в семействе методов (7.32) имеется значительное число свободных параметров. Основные соображения, положенные в основу этих версий, таковы: нужно получить формулы из семейства методов Рунге–Кутты (7.32), которые использовали бы одни и те же значения функции — правой части уравнения (7.1) — и определяли бы разные конкретные методы одного порядка (или смежных порядков, например, четвертого и пятого); при этом, чтобы по разности результатов подсчета приближенных значений решения по выведенным близким формулам (с одним и тем же шагом $h(!)$) можно было судить о точности одного из них.

Приведем один из таких методов, который называется *методом Кутты–Мерсона* или, иначе, *пятиэтапным методом Рунге–Кутты четвертого порядка*, а также *методом вложенных форм* [95].

На i -ом шаге решения задачи (7.1)–(7.2) последовательно вычисляют:

$$\begin{aligned}\eta_1^i &= f(x_i, y_i), \\ \eta_2^i &= f\left(x_i + \frac{h}{3}, y_i + \frac{h}{3}\eta_1^i\right), \\ \eta_3^i &= f\left(x_i + \frac{h}{3}, y_i + \frac{h}{6}\eta_1^i + \frac{h}{6}\eta_2^i\right), \\ \eta_4^i &= f\left(x_i + \frac{h}{2}, y_i + \frac{h}{8}\eta_1^i + \frac{3h}{8}\eta_2^i\right), \\ \eta_5^i &= f\left(x_i + h, y_i + \frac{h}{2}\eta_1^i - \frac{3h}{2}\eta_3^i + 2h\eta_4^i\right), \\ \tilde{y}_{i+1} &= y_i + \frac{h}{2}(\eta_1^i - 3\eta_3^i + 4\eta_4^i), \\ y_{i+1} &= y_i + \frac{h}{6}(\eta_1^i + 4\eta_4^i + \eta_5^i).\end{aligned}$$

После этого подсчитывают величину

$$R := 0.2|y_{i+1} - \tilde{y}_{i+1}|$$

и проводят сравнения. Если значение R окажется больше заданного допустимого уровня абсолютных погрешностей ε , то шаг

уменьшают вдвое $\left(h := \frac{h}{2}\right)$ и возвращаются к началу второго этапа, т.е. заново вычисляют η_2^i, η_3^i и т.д. Если $R \leq \varepsilon$, то считают $y(x_{i+1}) \approx y_{i+1}$ с точностью ε . При переходе к следующему шагу делается проверка на возможность увеличить расчетный шаг: если $R \leq \frac{\varepsilon}{64}$, то далее расчет ведется с шагом $h := 2h$.

Другие методы подобного типа можно найти, например, в монографии [4], где кроме приведенной модификации Мерсона метода Рунге–Кутты содержатся описание и расчетные формулы *модификаций Фельберга и Ингланда*.

Замечание 7.1. Осуществить *апостериорный контроль глобальной погрешности*, т.е. погрешности последнего вычисленного тем или иным шаговым методом значения $y_n \approx y(x_n) = y(b)$, можно, например, следующим образом [31].

Из интегрального равенства (7.3) при $x = x_n$ имеем

$$y(x_n) = y_0 + \int_{x_0}^{x_n} f(x, y(x)) dx. \quad (7.35)$$

Считая, что численным методом получена достаточно густая равномерная сеть приближенных значений y_i решения $y(x)$ задачи (7.1)–(7.2) и, следовательно, известны значения $f_i := f(x_i, y_i) \approx f(x_i, y(x_i))$, можно применить какую-либо квадратурную формулу замкнутого типа для вычисления фигурирующего в равенстве (7.35) интеграла (в частности, для этих целей можно использовать алгоритм Ромберга, см. гл.5). Величина разности между y_n и значением выражения $y_0 + \sum_i A_i f_i$, где A_i — коэффициенты используемой квадратурной формулы, позволяет приближенно судить о точности значения y_n . Оснований для этого тем больше, чем выше точность выполняемой для такого приближенного глобального контроля квадратуры.

УПРАЖНЕНИЯ

7.1. Последовательным дифференцированием уравнения

$$y' = xy + 1$$

найдите его приближенное частное решение, соответствующее начальному условию $y(0) = 0$, в виде отрезка степенного ряда до пятой степени включительно.

Можно ли получить тот же отрезок ряда методом последовательных приближений Пикара? Реализуйте пять итераций по формуле (7.5).

7.2. Запишите совокупность формул, определяющих метод пошаговых последовательных приближений, основанный на применении к обобщающему (7.3) интегральному уравнению

$$y(x) = y(x_i) + \int_{x_i}^x f(t, y(t)) dt$$

при каждом $x_i = x_0 + ih$ фиксированного числа k итераций по методу Пикара (за $y(x_i)$ в этом уравнении должно быть принято значение в точке x_i результата последней итерации).

Опробуйте такой метод на задаче предыдущего упражнения, варьируя h и k и добиваясь одинаковой точности результатов в точке $x = 2$, задавая эту точность самостоятельно.

7.3. Проверьте справедливость численных результатов, приведенных в таблицах примеров 7.1 и 7.2 (см. § 7.3 и § 7.7).

7.4. Убедитесь, что для линейного дифференциального уравнения с постоянными коэффициентами

$$y' = py + q \quad (7.36)$$

результат применения метода Рунге–Кутты второго порядка (7.30) не зависит от выбора значения параметра α ($\neq 0$).

Справедливо ли подобное утверждение для линейных ОДУ с переменными коэффициентами?

7.5. Можно ли утверждать, что метод Рунге–Кутты второго порядка (7.30), примененный к уравнению (7.36), приведет к тем же результатам, что и исправленный метод Эйлера (7.25)?

7.6. На задаче

$$y' = y^2 e^x - 2y, \quad y(0) = 0.5$$

протестируйте методы Рунге–Кутты второго порядка (7.30) при $\alpha = \frac{1}{2}$, $\alpha = 1$ и каком либо другом $\alpha \neq 0$, а также исправленный метод Эйлера (7.25), заполнив таблицу по образцу сводной таблицы результатов примеров 7.1, 7.2. Как ведут себя здесь поправки Ричардсона (7.34) сравнительно с истинными погрешностями? Насколько уточняются приближенные решения в результате прибавления поправок Ричардсона?

7.7. Разработайте какой-либо конкретный алгоритм для реального воплощения описанного в замечании 7.1 (см. § 7.7) способа приближенного апостериорного оценивания глобальной погрешности результатов численного интегрирования начальной задачи (7.1)–(7.2) методом Рунге–Кутты второго порядка.

Проверьте работу этого алгоритма в условиях упражнения 7.6 на отрезке $[0, 1]$.

7.8. Запишите вид функции $\varphi(x, y, h)$, при котором равенство (7.26) определяет классический метод Рунге–Кутты четвертого порядка.

ГЛАВА 8 || ЛИНЕЙНЫЕ МНОГОШАГОВЫЕ МЕТОДЫ

На интегро-интерполяционной основе выводится несколько семейств методов численного решения задачи Коши для ОДУ, имеющих перед многоэтапными одношаговыми методами таких же порядков точности то преимущество, что требуют меньших вычислительных затрат, благодаря использованию информации о решении в нескольких предыдущих точках. Показывается, как совместное применение явных и неявных формул одного порядка точности можно использовать для простого приближенного пошагового учета погрешностей. Наряду с конкретными семействами многошаговых методов, таких как методы Адамса и Милна, рассматривается общая структура многошаговых методов; демонстрируется возможность фиксирования параметров таких методов, исходя из понятия алгебраического порядка точности метода. Отмечается применимость всех рассматриваемых методов к решению систем дифференциальных уравнений и уравнений высших порядков. Описано построение методов Адамса и Адамса–Штёрмера для уравнений второго порядка.

§ 8.1. МНОГОШАГОВЫЕ МЕТОДЫ АДАМСА

Как и в предыдущей главе, будем строить численные методы решения начальной задачи

$$y' = f(x, y), \quad x \in [x_0, b], \quad (8.1)$$

$$y(x_0) = y_0. \quad (8.2)$$

Будем считать, что уже найдено несколько приближенных значений $y_j \approx y(x_j)$ ($j = 0, 1, \dots, i$) решения $y = y(x)$ задачи (8.1)–(8.2) на равномерной сетке $x_j = x_0 + jh$, и нужно получить правило для вычисления очередного значения $y_{i+1} \approx y(x_{i+1})$. Для вывода таких правил используем интегро-интерполяционный подход. А именно, проинтегрировав левую и правую части уравнения (8.1) по промежутку $[x_i, x_{i+1}]$, в полученном равенстве

$$y(x_{i+1}) = y(x_i) + \int_{x_i}^{x_{i+1}} f(x, y(x)) dx \quad (8.3)$$

под интеграл вместо функции $f(x, y(x))$ подставим интерполирующий ее многочлен $P_k(x)$. Хотя выражение функции $f(x, y(x))$, как функции одной переменной x , вообще говоря, неизвестно, ее дискретные приближенные значения $f(x_j, y_j) \approx f(x_j, y(x_j))$, обозначаемые в дальнейшем для краткости f_j , при $j = 1, 2, \dots, i$ можно считать известными. В таком случае, дополняя эти известные значения пока что неизвестным значением $f_{i+1} := f(x_{i+1}, y_{i+1})$, можно построить таблицу конечных разностей (табл.8.1), служащую основой для образования интерполяционных многочленов k -й степени для интерполирования назад из точек (x_i, f_i) и (x_{i+1}, f_{i+1}) .

Таблица 8.1

Таблица конечных разностей для построения
конечноразностных формул Адамса

x_j	f_j	Δf_j	$\Delta^2 f_j$	$\Delta^3 f_j$	...	$\Delta^k f_j$
x_{i-k}	f_{i-k}				...	
x_{i-k+1}	f_{i-k+1}	Δf_{i-k}				
x_{i-k+2}	f_{i-k+2}	Δf_{i-k+1}	$\Delta^2 f_{i-k}$	$\Delta^3 f_{i-k}$		
x_{i-k+3}	f_{i-k+3}	Δf_{i-k+2}	$\Delta^2 f_{i-k+1}$			
...	...	...	...	...		$\Delta^k f_{i-k}$
						$\Delta^k f_{i-k+1}$
x_{i-3}	f_{i-3}					
x_{i-2}	f_{i-2}	Δf_{i-3}				
x_{i-1}	f_{i-1}	Δf_{i-2}	$\Delta^2 f_{i-3}$	$\Delta^3 f_{i-3}$		
x_i	f_i	Δf_{i-1}	$\Delta^2 f_{i-2}$	$\Delta^3 f_{i-2}$		
x_i	f_i	Δf_i	$\Delta^2 f_{i-1}$			
x_{i+1}	f_{i+1}					

При интерполировании назад из узла x_i по второй интерполяционной формуле Ньютона (1.28) имеем

$$\begin{aligned}
 P_k(x) = P_k(x_i + qh) = & f_i + q\Delta f_{i-1} + \frac{q(q+1)}{2!} \Delta^2 f_{i-2} + \\
 & + \frac{q(q+1)(q+2)}{3!} \Delta^3 f_{i-3} + \dots + \frac{q(q+1)\dots(q+k-1)}{k!} \Delta^k f_{i-k} \quad (8.4)
 \end{aligned}$$

(см. конечные разности, подчеркнутые в табл.8.1 сплошной чертой), а из узла x_{i+1} по той же формуле получаем многочлен

$$\begin{aligned} \tilde{P}_k(x) = \tilde{P}_k(x_{i+1} + qh) = f_{i+1} + q\Delta f_i + \frac{q(q+1)}{2!}\Delta^2 f_{i-1} + \\ + \frac{q(q+1)(q+2)}{3!}\Delta^3 f_{i-2} + \dots + \frac{q(q+1)\dots(q+k-1)}{k!}\Delta^k f_{i-k+1} \end{aligned} \quad (8.5)$$

(использующий разности, подчеркнутые штриховой чертой).

Подстановка многочленов $P_k(x)$ и $\tilde{P}_k(x)$ в равенство (8.3) приводит к формулам для вычисления очередного значения $y_{i+1} \approx y(x_{i+1})$ вида

$$y_{i+1} = y_i + \int_{x_i}^{x_{i+1}} P_k(x) dx \quad (8.6)$$

и

$$y_{i+1} = y_i + \int_{x_i}^{x_{i+1}} \tilde{P}_k(x) dx. \quad (8.7)$$

В результате применения к интегралам в (8.6) и (8.7) формулы Ньютона–Лейбница получается два семейства методов (с параметром $k \in \mathbb{N}_0$), называемых **многошаговыми методами Адамса**^{*)}. Рассмотрим по отдельности каждое из этих семейств.

Экстраполяционные методы Адамса–Башфорта. Чтобы подставить в (8.6) многочлен (8.4), зависящий от переменной

$q = \frac{x - x_i}{h}$, сделаем в интеграле $\int_{x_i}^{x_{i+1}} P_k(x) dx$ замену переменной

$x = x_i + qh$, в соответствии с которой

$$\int_{x_i}^{x_{i+1}} P_k(x) dx = h \int_0^1 P_k(x_i + qh) dq.$$

Тогда формула (8.6) может быть переписана в виде

$$y_{i+1} = y_i + hI_k, \quad (8.8)$$

^{*)} Адамс Джон Кауч (1819–1892) — английский математик и астроном. Одновременно с Леверье им было предсказано существование планеты Нептун.

где

$$I_k := \int_0^1 P_k(x_i + qh) dq = \left[f_i q + \frac{q^2}{2} \Delta f_{i-1} + \left(\frac{q^3}{6} + \frac{q^2}{4} \right) \Delta^2 f_{i-2} + \right. \\ \left. + \frac{1}{6} \left(\frac{q^4}{4} + q^3 + q^2 \right) \Delta^3 f_{i-3} + \frac{1}{24} \left(\frac{q^5}{5} + \frac{3q^4}{2} + \frac{11q^3}{3} + 3q^2 \right) \Delta^4 f_{i-4} + \dots \right]_0^1 = \\ = f_i + \frac{1}{2} \Delta f_{i-1} + \frac{5}{12} \Delta^2 f_{i-2} + \frac{3}{8} \Delta^3 f_{i-3} + \frac{251}{720} \Delta^4 f_{i-4} + \dots \quad (8.9)$$

Таким образом, на основе (8.8) получается следующая конечно-разностная формула, определяющая **экстраполяционный метод Адамса–Башфорта**:

$$y_{i+1} = y_i + h \left(f_i + \frac{1}{2} \Delta f_{i-1} + \frac{5}{12} \Delta^2 f_{i-2} + \frac{3}{8} \Delta^3 f_{i-3} + \frac{251}{720} \Delta^4 f_{i-4} + \dots \right). \quad (8.10)$$

Посмотрим, что представляют собой наиболее простые частные случаи метода Адамса–Башфорта, соответствующие нескольким первым значениям параметра k в формуле (8.8). Сразу заметим, что при фиксировании $k = 0, 1, 2, \dots$ в (8.8) тем самым задается степень интерполяционного многочлена (нулевая, первая, вторая и т.д.) и, соответственно, число слагаемых, равное $1, 2, 3, \dots$, в правой части (8.9) (или, что то же, в скобках формулы (8.10)). Конечные разности в получающихся при этом конкретных формулах будем раскрывать через значения функции, приводя формулы к виду, называемому иногда **ординатным**. Имеем:

при $k = 0$

$$I_0 = f_i \Rightarrow y_{i+1} = y_i + hf(x_i, y_i); \quad (8.11)$$

при $k = 1$

$$I_1 = f_i + \frac{1}{2} \Delta f_{i-1} = \frac{3}{2} f_i - \frac{1}{2} f_{i-1} \Rightarrow \\ y_{i+1} = y_i + \frac{h}{2} [3f(x_i, y_i) - f(x_{i-1}, y_{i-1})]; \quad (8.12)$$

при $k = 2$

$$I_2 = I_1 + \frac{5}{12} \Delta^2 f_{i-2} = \frac{23}{12} f_i - \frac{16}{12} f_{i-1} + \frac{5}{12} f_{i-2} \Rightarrow$$

$$y_{i+1} = y_i + \frac{h}{12}[23f(x_i, y_i) - 16f(x_{i-1}, y_{i-1}) + 5f(x_{i-2}, y_{i-2})]; \quad (8.13)$$

при $k = 3$

$$I_3 = I_2 + \frac{3}{8}\Delta^3 f_{i-3} = \frac{55}{24}f_i - \frac{59}{24}f_{i-1} + \frac{37}{24}f_{i-2} - \frac{9}{24}f_{i-3} \Rightarrow$$

$$y_{i+1} = y_i + \frac{h}{24}[55f(x_i, y_i) - 59f(x_{i-1}, y_{i-1}) + 37f(x_{i-2}, y_{i-2}) - 9f(x_{i-3}, y_{i-3})]. \quad (8.14)$$

Формулы (8.11), (8.12), (8.13) и (8.14) определяют методы Адамса–Башфорта соответственно первого, второго, третьего и четвертого порядков. Относительно порядка метода (8.11) сомнений нет: мы узнаём *метод Эйлера* (7.8), вопрос о порядке которого обсуждался в § 7.2. Обоснование же утверждения о порядках остальных перечисленных конкретных методов состоит в следующем.

В общем случае, для $k + 1$ раз непрерывно дифференцируемой функции шаговая ошибка может быть получена сравнением равенства (8.6) с породившим его равенством (8.3), т.е. интегри-

рованием остаточного члена $R_k(x) = \frac{f^{(k+1)}(\xi, y(\xi))}{(k+1)!} \Pi_{k+1}(x)$ ин-

терполяционной формулы Лагранжа (см.(1.12)). Применительно к конечноразностной интерполяционной формуле (8.4) функция $R_k(x)$ преобразуется к виду

$$R_k(x_i + qh) = \frac{f^{(k+1)}(\xi, y(\xi))}{(k+1)!} q(q+1)\dots(q+k)h^{k+1}, \quad (8.15)$$

т.е. может считаться величиной $O(h^{k+1})$. Следовательно, локальная погрешность метода типа (8.8) будет составлять величину

$$h \int_0^1 R_k(x_i + qh) dq = O(h^{k+2}), \text{ а глобальная — величину } O(h^{k+1}).$$

Таким образом, метод Адамса, порождаемый интерполированием правой части уравнения (8.1) с помощью многочлена k -й степени, является методом $(k+1)$ -го порядка точности (относительно шага h).

Так как интерполяционный многочлен $P_k(x)$ степени k строился по $k+1$ значениям $f_{i-k}, f_{i-k+1}, \dots, f_{i-1}, f_i$, использующим, в свою очередь, $k+1$ значений (вообще говоря, приближенных) $y_{i-k}, y_{i-k+1}, \dots, y_{i-1}, y_i$ искомого решения $y(x)$, а количество этих значений (даже если какие-то значения между y_k и y_i явно не участвуют) определяет **шаговость метода**, то для методов семейства (8.10) мы наблюдаем совпадение порядка метода с тем, сколькошаговым он является. Следовательно, метод (8.12) — это двухшаговый метод Адамса–Башфорта второго порядка, метод (8.13) — трехшаговый метод Адамса–Башфорта третьего порядка, и т.д.

Название **экстраполяционный метод** связано с тем, что интерполяционный многочлен $P_k(x)$ для равенства (8.6) строился по узлам, расположенным на промежутке $[x_{i-k}, x_i]$, а применялся к отрезку $[x_i, x_{i+1}]$, т.е. производилось интерполирование в широком смысле, а в сущности делалось экстраполирование (по поводу нюансов этой терминологии см. сноски в § 1.2). Понятно, что в названии рассматриваемого далее второго семейства методов Адамса слово «интерполяционный» употребляется в узком смысле.

Интерполяционные методы Адамса–Моултона. В интеграле, фигурирующем в формуле (8.7), делаем замену $x = x_{i+1} + qh$ и подставляем в него выражение $\tilde{P}_k(x)$, определяемое формулой (8.5). Приходим к аналогичному (8.8) равенству

$$y_{i+1} = y_i + h\tilde{I}_k,$$

где

$$\begin{aligned} \tilde{I}_k &:= \int_{-1}^0 \tilde{P}_k(x_{i+1} + qh) dq = \left[f_{i+1}q + \frac{q^2}{2} \Delta f_i + \left(\frac{q^3}{6} + \frac{q^2}{4} \right) \Delta^2 f_{i-1} + \right. \\ &+ \frac{1}{6} \left(\frac{q^4}{4} + q^3 + q^2 \right) \Delta^3 f_{i-2} + \frac{1}{24} \left(\frac{q^5}{5} + \frac{3q^4}{2} + \frac{11q^3}{3} + 3q^2 \right) \Delta^4 f_{i-3} + \dots \left. \right]_{-1}^0 = \\ &= f_{i+1} - \frac{1}{2} \Delta f_i - \frac{1}{12} \Delta^2 f_{i-1} - \frac{1}{24} \Delta^3 f_{i-2} - \frac{19}{720} \Delta^4 f_{i-3} - \dots \quad (8.16) \end{aligned}$$

Отсюда следует конечноразностная формула интерполяционного

метода Адамса–Моултона

$$y_{i+1} = y_i + h \left[f_{i+1} - \frac{1}{2} \Delta f_i - \frac{1}{12} \Delta^2 f_{i-1} - \right. \\ \left. - \frac{1}{24} \Delta^3 f_{i-2} - \frac{19}{720} \Delta^4 f_{i-3} - \dots \right]. \quad (8.17)$$

Аналогично тому, как это делалось для методов Адамса–Башфорта, при $k = 0, 1, 2, 3$, т.е. фиксированием одного, двух, трех, четырех членов в представлении (8.16) интеграла $\tilde{I}_k(x)$, получаем следующие частные формулы:

при $k = 0$

$$y_{i+1} = y_i + hf(x_{i+1}, y_{i+1}), \quad (8.18)$$

при $k = 1$

$$y_{i+1} = y_i + \frac{h}{2} [f(x_{i+1}, y_{i+1}) + f(x_i, y_i)], \quad (8.19)$$

при $k = 2$

$$y_{i+1} = y_i + \frac{h}{12} [5f(x_{i+1}, y_{i+1}) + 8f(x_i, y_i) - f(x_{i-1}, y_{i-1})], \quad (8.20)$$

при $k = 3$

$$y_{i+1} = y_i + \frac{h}{24} [9f(x_{i+1}, y_{i+1}) + 19f(x_i, y_i) - \\ - 5f(x_{i-1}, y_{i-1}) + f(x_{i-2}, y_{i-2})]. \quad (8.21)$$

Формулы (8.18) и (8.19) определяют уже известные нам методы, а именно, **неявный метод Эйлера** (7.14) и **метод трапеций** (7.15), имеющие первый и второй порядки точности соответственно. Заметим, что оба эти метода являются одношаговыми, а следующие за ними методы Адамса–Моултона (8.20) и (8.21) третьего и четвертого порядков относятся, как легко видеть, соответственно к двухшаговым и трехшаговым методам. Таким образом, для интерполяционных методов Адамса–Моултона порядок шаговости на единицу ниже порядка точности метода (за тривиальным исключением, отвечающим случаю $k = 0$).

Важное различие в экстраполяционных и интерполяционных методах Адамса заключается в том, что первые из них являются явными, а вторые — неявными. Эти термины однозначно определяют, о каком из двух семейств методов Адамса идет

речь, а их сущность диктует особенности использования методов Адамса при практических расчетах, что найдет отражение в следующем параграфе.

§ 8.2. МЕТОДЫ ПРОГНОЗА И КОРРЕКЦИИ. ПРЕДИКТОР-КОРРЕКТОРНЫЕ МЕТОДЫ АДАМСА

Под названием *методы прогноза и коррекции* (иначе *методы предсказания и уточнения, предиктор-корректорные методы*) понимается совместное применение явных и неявных методов одинакового или смежных порядков. По явной формуле значение решения $y(x)$ задачи (8.1)–(8.2) в текущей (расчетной) точке x_{i+1} прогнозируется, т.е. находится его, быть может, достаточно грубое приближение, а с помощью неявной формулы, в правую часть которой подставляется спрогнозированное значение, оно уточняется (корректируется). Пример приближенного вычисления $y(x_{i+1})$ по такой явно-неявной схеме у нас уже есть: в § 7.3 рассматривалось парное использование явного метода Эйлера для предсказания и метода трапеций для уточнения (см. итерационную формулу (7.17), определяющую усовершенствованный метод Эйлера–Коши с итерационной обработкой).

Остановимся подробнее на методах прогноза и коррекции, базирующихся на парах явных и неявных методов Адамса одинакового порядка. Обозначим через y_{i+1}^B приближенное значение решения $y(x_{i+1})$, подсчитываемое по явной экстраполяционной формуле Адамса–Башфорта, и составим несколько пар из рассмотренных в предыдущем параграфе частных формул Адамса–Башфорта (8.11), (8.12), (8.13), (8.14) и Адамса–Моултона (8.18), (8.19), (8.20), (8.21).

Имеем следующие *предиктор-корректорные методы Адамса*:

первого порядка (он же *явно-неявный метод Эйлера*)

$$\begin{cases} y_{i+1}^B = y_i + hf(x_i, y_i), \\ y_{i+1} = y_i + hf(x_{i+1}, y_{i+1}^B); \end{cases}$$

второго порядка

$$\begin{cases} y_{i+1}^B = y_i + \frac{h}{2}[3f(x_i, y_i) - f(x_{i-1}, y_{i-1})], \\ y_{i+1} = y_i + \frac{h}{2}[f(x_{i+1}, y_{i+1}^B) + f(x_i, y_i)]; \end{cases}$$

третьего порядка

$$\begin{cases} y_{i+1}^B = y_i + \frac{h}{12}[23f(x_i, y_i) - 16f(x_{i-1}, y_{i-1}) + 5f(x_{i-2}, y_{i-2})], \\ y_{i+1} = y_i + \frac{h}{12}[5f(x_{i+1}, y_{i+1}^B) + 8f(x_i, y_i) - f(x_{i-1}, y_{i-1})]; \end{cases}$$

четвертого порядка

$$\begin{cases} y_{i+1}^B = y_i + \frac{h}{24}[55f(x_i, y_i) - 59f(x_{i-1}, y_{i-1}) + \\ \quad + 37f(x_{i-2}, y_{i-2}) - 9f(x_{i-3}, y_{i-3})], \\ y_{i+1} = y_i + \frac{h}{24}[9f(x_{i+1}, y_{i+1}^B) + 19f(x_i, y_i) - \\ \quad - 5f(x_{i-1}, y_{i-1}) + f(x_{i-2}, y_{i-2})]. \end{cases} \quad (8.22)$$

Одним из главных достоинств методов прогноза и коррекции является возможность контролировать шаговую погрешность сравнением двух полученных по явной и неявной формулам приближений к $y(x_{i+1})$. Покажем, как реализуется эта возможность для наиболее употребительного предиктор-корректорного метода Адамса четвертого порядка (8.22).

Вспомним, что первая из формул (8.22) была получена из общей формулы Адамса–Башфорта (8.10), а вторая — из общей формулы Адамса–Моултона (8.17), в которых последними брались разности третьего порядка (подынтегральная функция в равенстве (8.3) аппроксимировалась интерполяционным многочленом третьей степени). Считая, что расчетный шаг h достаточно мал и конечные разности с ростом их порядка убывают, главные части шаговых погрешностей формул Башфорта и Моултона четвертого порядка, в соответствии с (8.10) и (8.17), характеризуется величинами $\frac{251}{720}h\Delta^4 f_{i-4}$ для явной и $-\frac{19}{720}h\Delta^4 f_{i-3}$ для неявной формул. Таким образом, если наряду с введенным обозначением y_{i+1}^B обозначить через y_{i+1}^M приближенное значение $y(x_{i+1})$, получаемое по формуле Адамса–Моултона четвертого порядка, то можно записать два приближенных представления $y(x_{i+1})$:

$$y(x_{i+1}) \approx y_{i+1}^B + \frac{251}{720}h\Delta^4 f_{i-4} \quad (8.23)$$

и

$$y(x_{i+1}) \approx y_{i+1}^M - \frac{19}{720}h\Delta^4 f_{i-3}. \quad (8.24)$$

Отсюда видно, что если четвертые разности функции $f(x, y(x))$ в используемой части табл.8.1 конечных разностей практически постоянны (а это можно связать с удачным выбором величины шага h при достаточном запасе знаков в значениях $f(x_j, y_j)$), то, во-первых, значения y_{i+1}^B и y_{i+1}^M дают двусторонние приближения к точному решению $y(x_{i+1})$, а во-вторых, через разность между значениями y_{i+1}^B и y_{i+1}^M можно оценить точность каждого из них.

Действительно, приравнивая правые части приближенных равенств (8.23) и (8.24) и отождествляя $\Delta^4 f_{i-4}$ с $\Delta^4 f_{i-3}$, имеем:

$$y_{i+1}^M - y_{i+1}^B \approx \frac{19}{720} h \Delta^4 f_{i-3} + \frac{251}{720} h \Delta^4 f_{i-4} \approx \frac{3}{8} h \Delta^4 f_{i-3},$$

откуда

$$h \Delta^4 f_{i-3} \approx \frac{8}{3} (y_{i+1}^M - y_{i+1}^B).$$

Подставляя последнее в (8.24), получаем приближенное равенство

$$y(x_{i+1}) \approx y_{i+1}^M - \frac{19}{270} (y_{i+1}^M - y_{i+1}^B). \quad (8.25)$$

Использование приближенной формулы (8.25) может быть двояким. Переписав ее в виде

$$y(x_{i+1}) - y_{i+1}^M \approx -\frac{19}{270} (y_{i+1}^M - y_{i+1}^B),$$

применяем это для пошагового контроля точности:

если $\frac{19}{270} |y_{i+1}^M - y_{i+1}^B| < \varepsilon$, полагаем $y(x_{i+1}) \approx y_{i+1}^M$ с точностью ε и переходим к следующему шагу ($i := i+1$), иначе уменьшаем шаг h и снова подсчитываем y_{i+1}^M и y_{i+1}^B .

Другое назначение формулы (8.25) — это прямое применение ее правой части для получения уточненного значения:

полагаем

$$y_{i+1} \approx y_{i+1}^M - \frac{19}{270} (y_{i+1}^M - y_{i+1}^B). \quad (8.26)$$

Наверное, есть смысл контроль точности делать на каждом шаге, а к уточнению по формуле (8.26) прибегать при выводе окончательных результатов.

Замечание 8.1. При выводе формулы (8.25) под y_{i+1}^M мы понимаем значение, соответствующее «чистому» методу Адамса–Моултона четвертого порядка, т.е. y_{i+1}^M — это точная реализация неявной формулы (8.21). Вторая же формула предиктор-корректорного метода (8.22) соответствует лишь одному приближению к y_{i+1}^M по методу простых итераций, где в качестве начального приближения берется y_{i+1}^B . Поэтому применения формулы (8.25) к методу прогноза и коррекции (8.22) будут убедительны в том случае, если его вторая формула итерируется хотя бы один-два раза. Однако, чем больше таких итераций, тем ниже вычислительная эффективность этого метода, в целом весьма высокая по сравнению с многоэтапными методами Рунге–Кутты.

§ 8.3. МЕТОД МИЛНА ЧЕТВЕРТОГО ПОРЯДКА

Рассмотрим еще один широко известный метод прогноза и коррекции — *метод Милна*.

Для вывода первой формулы Милна (т.е. формулы предсказания) проинтегрируем данное уравнение (8.1) на промежутке $[x_{i-3}, x_{i+1}]$ и в полученном интегральном равенстве

$$y(x_{i+1}) = y(x_{i-3}) + \int_{x_{i-3}}^{x_{i+1}} f(x, y(x)) dx \quad (8.27)$$

подынтегральную функцию $f(x, y(x))$ заменим первым интерполяционным многочленом Ньютона $P_3(x)$, построенным по четырем узлам $x_{i-3}, x_{i-2}, x_{i-1}, x_i$ с предполагающимися уже известными приближенными значениями

$$f_{i-3} := f(x_{i-3}, y_{i-3}) \approx f(x_{i-3}, y(x_{i-3})),$$

$$f_{i-2} := f(x_{i-2}, y_{i-2}) \approx f(x_{i-2}, y(x_{i-2})),$$

$$f_{i-1} := f(x_{i-1}, y_{i-1}) \approx f(x_{i-1}, y(x_{i-1})),$$

$$f_i := f(x_i, y_i) \approx f(x_i, y(x_i)).$$

Тогда, после замены переменной $x \approx x_{i-3} + qh$, на основании (8.27) имеем:

$$\begin{aligned} y(x_{i+1}) \approx y_{i-3} + h \int_0^4 P_3(x_{i-3} + qh) dq = y_{i-3} + h \int_0^4 [f_{i-3} + q\Delta f_{i-3} + \\ + \frac{q(q-1)}{2!} \Delta^2 f_{i-3} + \frac{q(q-1)(q-2)}{3!} \Delta^3 f_{i-3}] dq = y_{i-3} + \end{aligned}$$

$$+ h \left[f_{i-3} q + \frac{q^2}{2} \Delta f_{i-3} + \frac{1}{2} \left(\frac{q^3}{3} - \frac{q^2}{2} \right) \Delta^2 f_{i-3} + \frac{1}{6} \left(\frac{q^4}{4} - q^3 + q^2 \right) \Delta^3 f_{i-3} \right]_0^4 =$$

$$= y_{i-3} + \frac{4}{3} h (3 f_{i-3} + 6 \Delta f_{i-3} + 5 \Delta^2 f_{i-3} + 2 \Delta^3 f_{i-3}).$$

Отсюда, выразив конечные разности через значения функции, получаем **первую формулу Милна** (предсказания)

$$\hat{y}_{i+1} = y_{i-3} + \frac{4}{3} h [2 f(x_i, y_i) - f(x_{i-1}, y_{i-1}) + 2 f(x_{i-2}, y_{i-2})], \quad (8.28)$$

которую, очевидно, следует отнести к экстраполяционным.

Главный член локальной погрешности формулы (8.28) находим интегрированием следующего (первого из неучтенных) слагаемого интерполяционного многочлена Ньютона. Именно:

$$y(x_{i+1}) - \hat{y}_{i+1} \approx h \int_0^4 \frac{q(q-1)(q-2)(q-3)}{4!} \Delta^4 f_{i-3} dq =$$

$$= \frac{h}{24} \Delta^4 f_{i-3} \int_0^4 (q^4 - 6q^3 + 11q^2 - 6q) dq = \frac{14}{45} h \Delta^4 f_{i-3}.$$

Считая четвертые разности примерно одинаковыми, опустим индекс у функции f в записи $\Delta^4 f_{i-3}$; в результате получаем следующее приближенное представление решения в точке x_{i+1} :

$$y(x_{i+1}) \approx \hat{y}_{i+1} + \frac{14}{45} h \Delta^4 f. \quad (8.29)$$

Вывод второй формулы Милна более прост. Проинтегрируем уравнение (8.1) теперь на промежутке $[x_{i-1}, x_{i+1}]$ и в полученном равенстве

$$y(x_{i+1}) = y(x_{i-1}) + \int_{x_{i-1}}^{x_{i+1}} f(x, y(x)) dx$$

применим к интегралу простейшую формулу Симпсона (5.30). Имеем

$$y(x_{i+1}) = y(x_{i-1}) + \frac{h}{3} [f(x_{i+1}, y(x_{i+1})) + 4 f(x_i, y(x_i)) +$$

$$+ f(x_{i-1}, y(x_{i-1}))] - \frac{h^5}{90} f^{IV}(\xi_i). \quad (8.30)$$

Отбрасывая здесь остаточный член и заменяя значения решения

$y(x_{i-1})$ и $y(x_i)$ известными приближенными значениями y_{i-1} и y_i , а стоящее в правой части под знаком функции f неизвестное значение $y(x_{i+1})$ тем значением $\hat{y}_{i+1}$, которое получается в результате вычислений по явной первой формуле Милна (8.28), приходим ко **второй формуле Милна** (уточнения)

$$y_{i+1} = y_{i-1} + \frac{h}{3}[f(x_{i+1}, \hat{y}_{i+1}) + 4f(x_i, y_i) + f(x_{i-1}, y_{i-1})], \quad (8.31)$$

являющейся интерполяционной.

Для вывода приближенной оценки шаговой погрешности воспользуемся приближенным равенством $f^{IV}(\xi) \approx \frac{\Delta^4 f}{h^4}$, где

$\Delta^4 f$ так же, как и в (8.29), — условная запись практически постоянных четвертых разностей. Исходя из точного равенства (8.30), локальную погрешность получаемого с помощью формулы (8.31) (возможно, с итерационной обработкой, см. замечание 8.1) приближенного значения y_{i+1} можно приближенно

охарактеризовать величиной $-\frac{h}{90}\Delta^4 f$, т.е.

$$y(x_{i+1}) \approx y_{i+1} - \frac{h}{90}\Delta^4 f. \quad (8.32)$$

Сравнение (8.29) и (8.32) дает:

$$y_{i+1} - \hat{y}_{i+1} \approx 29\frac{h}{90}\Delta^4 f \Rightarrow \frac{h}{90}\Delta^4 f \approx \frac{y_{i+1} - \hat{y}_{i+1}}{29},$$

следовательно,

$$y(x_{i+1}) - y_{i+1} \approx \frac{\hat{y}_{i+1} - y_{i+1}}{29}. \quad (8.33)$$

Таким образом, при численном интегрировании начальной задачи (8.1)–(8.2) методом Милна четвертого порядка, определенным формулами (8.28) и (8.31), на каждом i -м шаге следует вычислять величину

$$d_{i+1} := \frac{\hat{y}_{i+1} - y_{i+1}}{29}$$

и сравнивать ее модуль с величиной $\varepsilon > 0$ допустимой шаговой

погрешности. Если $|d_{i+1}| < \varepsilon$, то за $y(x_{i+1})$ принимается полученное по второй формуле Милна значение y_{i+1} (или его уточненное значение $y_{i+1} := y_{i+1} + d_{i+1}$); иначе шаг должен быть уменьшен.

Фигурирующая в приближенном равенстве (8.33) постоянная $\frac{1}{29}$ примерно вдвое меньше постоянной $\frac{19}{270} \approx \frac{1}{14}$ в аналогичном равенстве (8.25) для предиктор-корректорного метода Адамса четвертого порядка (8.22), что характеризует метод Милна как несколько более точный при одинаковых вычислительных затратах. Но в дальнейшем (см. гл. 9, в частности, § 9.5) появятся и другие критерии для сравнения этих конкурирующих многошаговых методов.

§ 8.4. ОБЩИЙ ВИД ЛИНЕЙНЫХ МНОГОШАГОВЫХ МЕТОДОВ. УСЛОВИЯ СОГЛАСОВАННОСТИ

Все m -шаговые методы Адамса можно описать одной формулой

$$y_{i+1} = y_i + h \sum_{j=0}^m \beta_j f(x_{i+1-j}, y_{i+1-j}), \quad (8.34)$$

где m должно быть фиксированным натуральным числом, а i может принимать значения $m-1, m, m+1, \dots$. При $\beta_0 = 0$ формула (8.34) определяет явные, а при $\beta_0 \neq 0$ — неявные методы, которые с единственными конкретными наборами коэффициентов $\beta_1, \beta_2, \beta_3, \dots, \beta_m$ в первом и $\beta_0, \beta_1, \beta_2, \dots, \beta_m$ во втором случаях являются соответственно экстраполяционными и интерполяционными методами Адамса. Как известно, m -шаговый метод Адамса при достаточной гладкости решения имеет m -й порядок точности. Так как при любом его порядке для реализации одного шага требуется вычисление лишь одного нового значения функции (или двух при парном применении явных и неявных методов Адамса, см. предыдущий параграф), то при построении решения на большом промежутке выгодно применять методы Адамса достаточно высоких порядков. Но при этом возникают проблемы с вычислением первых $m-1$ «разгонных» значений $y_1, y_2, y_3, \dots, y_{m-1}$ (учитываем, что значение y_0 задано). Для их получения разработаны специальные процедуры разгона [4, 31]. Впрочем, можно обойтись применением для этих целей одношаговых методов Рунге–Кутты.

Аналогично (8.34) при $m = 4$ выглядят и формулы Милна, только в формуле прогноза (8.28) первым слагаемым стоит y_{i-3} , а в формуле коррекции (8.31) — y_{i-1} . Похожий на (8.34) вид

имеют также многошаговые методы Нистрёма, Хемминга и др. (см. [4]). Все эти методы можно описать формулой

$$y_{i+1} = \sum_{j=1}^m \alpha_j y_{i+1-j} + h \sum_{j=0}^m \beta_j f(x_{i+1-j}, y_{i+1-j}), \quad (8.35)$$

которая задает **общий вид линейных многошаговых методов** (приближенное значение решения $y(x)$ в точке x_{i+1} представляется в виде линейной комбинации нескольких приближенных значений решения и его производной в этой и в предшествующих m точках).

К построению конкретных многошаговых методов, т.е. к фиксированию параметров α_j , β_j в формуле (8.35) при фиксированных значениях $m = 1, 2, 3, \dots$ можно подойти следующим, отличным от рассмотренных ранее способом.

Обратившись к любой из выведенных в двух предыдущих параграфах многошаговых формул вычисления приближенного значения y_{i+1} решения $y(x_{i+1})$ уравнения $y' = f(x, y)$, мы видим, что локальная ошибка, точнее, ее главная часть, представляется в виде (см. (8.23), (8.24), (8.29), (8.32))

$$y(x_{i+1}) - y_{i+1} = Ch\Delta^k f = C_1 h^{k+1} f^{(k)}(\xi_i) = C_1 h^{k+1} y^{(k+1)}(\xi_i), \quad (8.36)$$

где C и C_1 — некоторые постоянные, ξ_i — некоторая точка, а k — порядок метода. Так как многочлен k -й степени имеет нулевую $(k+1)$ -ю производную, то, значит, метод получения y_{i+1} , для которого имеет место (8.36), точен, если его решение $y(x)$ — многочлен k -й (и менее) степени.

Подмеченный факт приводит к мысли положить в основу построения методов численного интегрирования задачи Коши (8.1)–(8.2) **алгебраический порядок точности**, под которым здесь понимается максимальная степень k многочлена $y(x) = P_k(x)$, обращающего в нуль погрешность (8.36).

Будем подставлять в формулу (8.35), определяющую общий вид m -шаговых методов, в роли решения $y(x)$ представителей многочленов $P_k(x)$ — степенные функции $1, x, x^2, \dots, x^k$, и будем считать, что такая подстановка вместо y_j значений $y(x_j)$ не порождает погрешностей, т.е. получающиеся при этом равенства являются точными.

Пусть $k = 0$, т.е. возьмем $y(x) \equiv 1$. Тогда, так как $y(x_j) = 1$ при любом значении j , а $y' = f(x, y) \equiv 0$, то (8.35) в этом случае превращается в равенство

$$1 = \sum_{j=1}^m \alpha_j \cdot 1 - h \sum_{j=0}^m \beta_j \cdot 0.$$

Следовательно, параметры α_j любого из методов семейства (8.35) должны подчиняться условию

$$\sum_{j=1}^m \alpha_j = 1. \quad (8.37)$$

Положим теперь $k = 1$ и будем считать решением $y(x) \equiv x$. В силу $y' = f(x, y) \equiv 1$, результат подстановки такой функции в (8.35) приведет к равенству

$$x_{i+1} = \sum_{j=1}^m \alpha_j x_{i+1-j} + h \sum_{j=0}^m \beta_j, \quad (8.38)$$

которое должно быть справедливым при любых $i = m-1, m, m+1, \dots$. Взяв $i = m$, исходя из выражения i -го узла через шаг $x_i = x_0 + ih$, от (8.38) приходим к равенству

$$x_0 + (m+1)h = \sum_{j=1}^m \alpha_j [x_0 + (m+1-j)h] + h \sum_{j=0}^m \beta_j.$$

Переписав его в виде

$$x_0 + (m+1)h = x_0 \sum_{j=1}^m \alpha_j + (m+1)h \sum_{j=1}^m \alpha_j - h \sum_{j=1}^m j \alpha_j + h \sum_{j=0}^m \beta_j,$$

учитываем найденное выше условие (8.37), в результате чего получаем простое требование ко второй группе параметров семейства m -шаговых методов (8.35):

$$\sum_{j=0}^m \beta_j = \sum_{j=1}^m j \alpha_j. \quad (8.39)$$

Совокупность требований (8.37), (8.39) называют **условиями согласованности параметров** линейных многошаговых методов (8.35). Нами они выведены как необходимые условия того, чтобы произвольный метод этого семейства имел, по меньшей мере, первый алгебраический порядок точности. Доказано [89], что они являются и достаточными для этого условиями.

Посмотрим, что представляют собой одношаговые методы из семейства m -шаговых методов, т.е. зафиксируем в (8.35) $m = 1$. Имеем трехпараметрическую формулу

$$y_{i+1} = \alpha_1 y_i + h(\beta_0 f_{i+1} + \beta_1 f_i). \quad (8.40)$$

Чтобы она определяла метод не ниже первого порядка, согласно условиям (8.37) и (8.39), должно быть

$$\alpha_1 = 1, \quad \beta_0 + \beta_1 = 1.$$

Таким образом, остается лишь одна степень свободы, т.е. на самом деле, (8.40) — это однопараметрическое семейство одношаговых методов

$$y_{i+1} = y_i + h[\beta_0 f_{i+1} + (1 - \beta_0) f_i]. \quad (8.41)$$

При любом значении β_0 метод (8.41) точен для многочленов первой степени. Например, при $\beta_0 = 0$ — это **явный метод Эйлера**, а при $\beta_0 = 1$ — **неявный метод Эйлера**. Оба они точны, если решением задачи (8.1)–(8.2) является линейная функция.

Попробуем подобрать параметр β_0 так, чтобы формула (8.41) была точной для многочленов второй степени, в частности, для функции $y(x) \equiv x^2$. Для этой функции $y' = f(x, y(x)) = 2x$, и через i -й узел и шаг h равенство (8.41) записывается так:

$$(x_i + h)^2 = x_i^2 + h[\beta_0 2(x_i + h) + (1 - \beta_0) 2x_i].$$

После элементарных упрощений в последнем равенстве приходим к значению параметра $\beta_0 = \frac{1}{2}$, подстановка которого в (8.41) приводит к методу

$$y_{i+1} = y_i + \frac{h}{2}(f_{i+1} + f_i).$$

В нем мы узнаём **метод трапеций** (7.15), единственный из семейства методов (8.41) (иначе, единственный из одношаговых методов семейства (8.35)), имеющий второй порядок точности.

Пусть теперь в общей формуле (8.35) $m = 2$, т.е. будем рассматривать всевозможные **двухшаговые методы** этого семейства:

$$y_{i+1} = \alpha_1 y_i + \alpha_2 y_{i-1} + h(\beta_0 f_{i+1} + \beta_1 f_i + \beta_2 f_{i-1}). \quad (8.42)$$

Потребуем, чтобы такой метод имел, по меньшей мере, второй порядок точности, т.е. чтобы подстановка в (8.42) вместо y_j значений $y(x_j)$ для $y(x) = x^2$ оставляла это равенство точным. Учтем, что оно должно быть верным при любом $i = 1, 2, 3, \dots$, и что, в силу однородности функции — предполагаемого решения $y(x) = x^2$, можно считать $i = 1$ и, соответственно, $x_0 = 0$, $x_1 = h$, $x_2 = 2h$. В этих точках решение $y = x^2$ принимает значения $0, h^2, 4h^2$, а его производная, т.е. функция $f(x, y(x)) = 2x$ — значения $0, 2h, 4h$. Следовательно, при $i = 1$ равенство (8.42) приобретает вид

$$4h^2 = \alpha_1 h^2 + 4\beta_0 h^2 + 2\beta_1 h^2.$$

Записав для этого случая еще условия согласованности (8.37), (8.39), получаем систему трех линейных уравнений с пятью неизвестными параметрами:

$$\begin{cases} \alpha_1 + \alpha_2 = 1, \\ \beta_0 + \beta_1 + \beta_2 = \alpha_1 + 2\alpha_2, \\ \alpha_1 + 4\beta_0 + 2\beta_1 = 4. \end{cases} \quad (8.43)$$

Будем считать свободными параметры

$$\alpha := \alpha_1 \quad \text{и} \quad \beta := \beta_0.$$

Тогда из системы (8.43) находим выражения остальных параметров:

$$\alpha_2 = 1 - \alpha, \quad \beta_1 = 2 - 2\beta - \frac{\alpha}{2}, \quad \beta_2 = \beta - \frac{\alpha}{2}.$$

Их подстановка в (8.42) приводит к двухпараметрическому семейству двухшаговых методов, имеющих, по крайней мере, второй алгебраический порядок точности:

$$y_{i+1} = \alpha y_i + (1 - \alpha)y_{i-1} + h \left[\beta f_{i+1} + \left(2 - 2\beta - \frac{\alpha}{2} \right) f_i + \left(\beta - \frac{\alpha}{2} \right) f_{i-1} \right]. \quad (8.44)$$

Легко видеть, что при $\alpha = 1$, $\beta = 0$ это семейство содержит частным случаем метод Адамса–Башфорта второго порядка (см.(8.12)), а при $\alpha = 1$, $\beta = \frac{1}{2}$ — метод Адамса–Моултона второго порядка (8.19) (он же — метод трапеций, единственный из содержащихся в (8.44) одношаговых методов).

Выберем еще одну степень свободы, потребовав, чтобы метод (8.44) имел третий порядок. Аналогично тому, как получали третье уравнение системы (8.43), полагаем $y(x) \equiv x^3$ и используем значения этой функции $0, h^3, 8h^3$ и значения ее производной $0, 3h^2, 12h^2$ на сетке $x_0 = 0$, $x_1 = h$, $x_2 = 2h$ в равенстве (8.44) с фиксированным $i = 1$. Имеем:

$$8h^3 = \alpha h^3 + 12\beta h^3 + \left(6 - 6\beta - \frac{3\alpha}{2} \right) h^3,$$

откуда получаем выражение $\beta = \frac{1}{3} + \frac{\alpha}{12}$. Таким образом, двухпараметрическое семейство методов, по крайней мере, второго

порядка (8.44) превращается в однопараметрическое семейство методов, по крайней мере, третьего порядка

$$y_{i+1} = \alpha y_i + (1 - \alpha) y_{i-1} + h \left[\left(\frac{1}{3} + \frac{\alpha}{12} \right) f_{i+1} + \left(\frac{4}{3} - \frac{2\alpha}{3} \right) f_i + \left(\frac{1}{3} - \frac{5\alpha}{12} \right) f_{i-1} \right]. \quad (8.45)$$

Распоряжаться оставшимся единственным параметром α в (8.45) можно из разных соображений. Например, чтобы формула (8.45) была явной, нужно обнулить коэффициент при f_{i+1} , т.е. положить $\alpha = -4$. В результате этого приходим к двухшаговому методу третьего порядка (не встречавшемуся ранее)

$$y_{i+1} = 5 y_{i-1} - 4 y_i + h(4 f_i + 2 f_{i-1}). \quad (8.46)$$

Чтобы метод (8.45) относился к семейству (8.34) методов Адамса, следует обнулить коэффициент при y_{i-1} , т.е. взять $\alpha = 1$. При этом получаем известный метод Адамса–Моултона третьего порядка (8.20). Легко подметить, что при $\alpha = 0$ формула (8.45) определяет метод Симпсона, иначе, вторую формулу Милна (8.31), имеющую, как известно, четвертый порядок точности (см. выражение ее остаточного члена, точнее, локальной ошибки в формуле (8.30)).

Как будет выяснено позже, свободный параметр в многошаговых методах может потребоваться для того, чтобы за его счет улучшить численную устойчивость метода.

Замечание 8.2. Существуют линейные многошаговые методы, не вписывающиеся в общую формулу (8.35). К таким относятся так называемые *методы с забеганием вперед*. В основе вывода этих методов тоже лежит интегро-интерполяционный подход, но, в отличие от методов Адамса, опирающихся на второй интерполяционный многочлен Ньютона при аппроксимации подынтегральной функции f в интегральном равенстве (8.3), здесь в том же равенстве (8.3) применяется интерполяционный многочлен Бесселя (1.33). Он использует более близкую к промежутку $[x_i, x_{i+1}]$ информацию о функции, что обеспечивает большую локальную точность. Однако в результате получаются формулы, в которых участвуют недоступные для непосредственного (и даже неявного) вычисления значения, следующие за y_{i+1} . Например, основная формула наиболее известного представителя методов с забеганием вперед — *метода Коуэлла четвертого порядка* — имеет вид

$$y_{i+1} = y_i + \frac{h}{24} [-f_{i+2} + 13f_{i+1} + 13f_i - f_{i-1}]. \quad (8.47)$$

Ясно, что одиночное применение таких формул невозможно, и методы с

Введем следующие векторные обозначения:

$$Y := \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad Y' := \begin{pmatrix} y'_1 \\ y'_2 \\ \vdots \\ y'_n \end{pmatrix}, \quad F(x, Y) := \begin{pmatrix} f_1(x, y_1, \dots, y_n) \\ f_2(x, y_1, \dots, y_n) \\ \dots \dots \dots \dots \dots \dots \\ f_n(x, y_1, \dots, y_n) \end{pmatrix}, \quad Y_0 := \begin{pmatrix} y_1^0 \\ y_2^0 \\ \vdots \\ y_n^0 \end{pmatrix}.$$

С ними данная задача Коши принимает вид

$$Y' = F(x, Y), \quad Y(x_0) = Y_0, \quad (8.48)$$

по форме точно такой же, как и рассматриваемая до сих пор задача (8.1)–(8.2).

Легко понять, что к векторному дифференциальному уравнению (8.48), в принципе, можно применить любой из численных методов, изучавшихся в этой и в предыдущей главах (благодаря линейной структуре методов).

При этом скалярными величинами в формулах, определяющих методы, являются только независимая переменная x и расчетный шаг h ; всем остальным величинам соответствуют введенные выше векторы размерности n . Следует лишь учесть, что при контроле точности вместо модуля нужно использовать норму вектора (например, норму-максимум).

Пример 8.1. Дана начальная задача

$$\begin{cases} y' = e^{-y^2 - z^2} + 2x, \\ z' = 2y^2 + z, \end{cases} \quad \begin{cases} y(0) = 0.5, \\ z(0) = 1. \end{cases}$$

Сделаем три шага явно-неявным методом Эйлера (предиктор-корректорным методом Адамса первого порядка), полагая $h = 0.1$. Предварительно для размерности $n = 2$ выполним покомпонентную запись расчетных формул. Для данной системы они выглядят так:

$$\begin{cases} \tilde{y}_{i+1} = y_i + 0.1(e^{-y_i^2 - z_i^2} + 2x_i), \\ \tilde{z}_{i+1} = z_i + 0.1(2y_i^2 + z_i); \end{cases}$$

$$\begin{cases} y_{i+1} = y_i + 0.1(e^{-\tilde{y}_{i+1}^2 - \tilde{z}_{i+1}^2} + 2x_i + 0.2), \\ z_{i+1} = z_i + 0.1(2\tilde{y}_{i+1}^2 + \tilde{z}_{i+1}). \end{cases}$$

Результаты счета с небольшой точностью по этим формулам при $i = 0, 1, 2$ представлены следующей таблицей.

i	x_{i+1}	$\tilde{y}_{i+1}$	$\tilde{z}_{i+1}$	y_{i+1}	z_{i+1}	$\max \{ y_{i+1} - \tilde{y}_{i+1} , z_{i+1} - \tilde{z}_{i+1} \}$
0	0.1	≈ 0.5287	1.15	≈ 0.5401	≈ 1.1709	≈ 0.02
1	0.2	≈ 0.5791	≈ 1.3463	≈ 0.5918	≈ 1.3726	≈ 0.03
2	0.3	≈ 0.6425	≈ 1.5799	≈ 0.6573	≈ 1.6132	≈ 0.03

§ 8.6. ЧИСЛЕННОЕ РЕШЕНИЕ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ ВЫСШИХ ПОРЯДКОВ. МЕТОДЫ АДАМСА–ШТЁРМЕРА

Одним из основных способов численного решения начальных задач для дифференциальных уравнений высших порядков является сведение их к соответствующим задачам для систем уравнений первого порядка.

Будем рассматривать уравнение второго порядка

$$y'' = f(x, y, y'), \quad x \in [x_0, b] \tag{8.49}$$

с начальными условиями

$$y(x_0) = y_0, \quad y'(x_0) = y'_0. \tag{8.50}$$

Введя новую переменную равенством $z = y'$, от уравнения (8.49) переходим к эквивалентной ему системе

$$\begin{cases} y' = z, \\ z' = f(x, y, z). \end{cases} \tag{8.51}$$

Начальные условия (8.50) для нее перепишутся в виде

$$\begin{cases} y(x_0) = y_0, \\ z(x_0) = y'_0. \end{cases} \tag{8.52}$$

К задаче Коши (8.51)–(8.52), в соответствии со сказанным в предыдущем параграфе, можно применить любой численный процесс из рассмотренных.

Альтернативой способу сведения к системам могут послужить методы, выводимые специально для уравнений высших порядков, например, на интегро-интерполяционной основе. Покажем, как может быть построен такой метод типа метода Адамса для задачи (8.49)–(8.50).

Будем считать, что уже известны значения $y_i \approx y(x_i)$, $y'_i \approx y'(x_i)$ и несколько предыдущих приближенных значений решения $y(x)$ и его производной $y'(x)$. Требуется получить

формулы для вычисления величин y_{i+1} и y'_{i+1} — приближенных значений функций $y(x)$ и $y'(x)$ в очередной расчетной точке $x_{i+1} = x_i + h$. Для их вывода проинтегрируем уравнение (8.49) на промежутке $[x_i, x]$, в результате чего приходим к интегральному уравнению

$$y'(x) = y'(x_i) + \int_{x_i}^x f(t, y, y') dt \quad (8.53)$$

относительно неизвестной функции $y'(x)$. Зафиксируем в нем $x = x_{i+1}$, заменим $y'(x_i)$ известным приближенным значением y'_i

и подставим в интеграле $\int_{x_i}^{x_{i+1}} f(x, y(x), y'(x)) dx$ вместо функции $f(x, y(x), y'(x))$ интерполяционный многочлен Ньютона (второй)

$$P_k(x) = P_k(x_i + qh) = f_i + q\Delta f_{i-1} + \frac{q(q+1)}{2!} \Delta^2 f_{i-2} + \dots \\ \dots + \frac{q(q+1)\dots(q+k-1)}{k!} \Delta^k f_{i-k}, \quad (8.54)$$

построенный по значениям $f_j := f(x_j, y_j, y'_j)$ (где $j = i-k, i-k+1, \dots, i-1, i$). После интегрирования получаем конечноразностную формулу

$$y'_{i+1} = y'_i + h \left(f_i + \frac{1}{2} \Delta f_{i-1} + \frac{5}{12} \Delta^2 f_{i-2} + \right. \\ \left. + \frac{3}{8} \Delta^3 f_{i-3} + \frac{251}{720} \Delta^4 f_{i-4} + \dots \right), \quad (8.55)$$

которая представляет собой экстраполяционный метод Адамса-Башфорта (см. (8.10)), применяемый здесь для нахождения приближенных значений производной, а не самого решения.

Для выражения значения $y_{i+1} \approx y(x_{i+1})$ решения данной задачи через те же конечные разности, что участвуют в формуле (8.55), проинтегрируем равенство (8.53) в пределах от x_i до x_{i+1} . Имеем точное равенство

$$y(x_{i+1}) = y(x_i) + \int_{x_i}^{x_{i+1}} [y'(x_i) + \int_{x_i}^x f(t, y, y') dt] dx,$$

которое заменяем приближенным равенством

$$y(x_{i+1}) \approx y_i + hy'_i + \int_{x_i}^{x_{i+1}} \int_{x_i}^x f(t, y, y') dt dx. \quad (8.56)$$

Результат подстановки в правую часть последнего вместо $f(t, y(t), y'(t))$ интерполяционного многочлена $P_k(t)$ вида (8.54) принимаем за искомое значение y_{i+1} , т.е. полагаем

$$y_{i+1} = y_i + hy'_i + \int_{x_i}^{x_{i+1}} \int_{x_i}^x P_k(t) dt dx. \quad (8.57)$$

Сделав замены переменных интегрирования $x = x_i + ph$ и $t = x_i + qh$, выполним двойное интегрирование многочлена:

$$\begin{aligned} \int_{x_i}^{x_{i+1}} \int_{x_i}^x P_k(t) dt dx &= h^2 \int_0^1 dp \int_0^p P_k(x_i + qh) dq = \\ &= h^2 \int_0^1 dp \int_0^p \left[f_i + q\Delta f_{i-1} + \frac{q(q+1)}{2!} \Delta^2 f_{i-2} + \frac{q(q+1)(q+2)}{3!} \Delta^3 f_{i-3} + \right. \\ &\quad \left. + \frac{q(q+1)(q+2)(q+3)}{4!} \Delta^4 f_{i-4} + \dots \right] dq = \\ &= h^2 \int_0^1 \left[f_i p + \frac{p^2}{2} \Delta f_{i-1} + \left(\frac{p^3}{6} + \frac{p^2}{4} \right) \Delta^2 f_{i-2} + \frac{1}{6} \left(\frac{p^4}{4} + p^3 + p^2 \right) \Delta^3 f_{i-3} + \right. \\ &\quad \left. + \frac{1}{24} \left(\frac{p^5}{5} + \frac{3p^4}{2} + \frac{11p^3}{3} + 3p^2 \right) \Delta^4 f_{i-4} + \dots \right] dp = \\ &= h^2 \left(\frac{1}{2} f_i + \frac{1}{6} \Delta f_{i-1} + \frac{1}{8} \Delta^2 f_{i-2} + \frac{19}{180} \Delta^3 f_{i-3} + \frac{3}{32} \Delta^4 f_{i-4} + \dots \right). \end{aligned}$$

Подставляя это выражение двойного интеграла в (8.57), получаем конечноразностную формулу для приближенного вычисления самого решения $y(x)$ в точке x_{i+1} :

$$y_{i+1} = y_i + hy'_i + \frac{h^2}{2} \left(f_i + \frac{1}{3} \Delta f_{i-1} + \frac{1}{4} \Delta^2 f_{i-2} + \frac{19}{90} \Delta^3 f_{i-3} + \frac{3}{16} \Delta^4 f_{i-4} + \dots \right). \quad (8.58)$$

Совокупность формул (8.55) и (8.58) определяют для задачи (8.49)–(8.50) семейство многошаговых экстраполяционных методов Адамса в конечноразностном виде. Фиксируя в них порядок последней используемой разности, тем самым задаем шаговость

и порядок метода^{*}), причем первые из отбрасываемых слагаемых (с учетом множителя за скобками) грубо характеризуют шаговую погрешность вычисления y'_{i+1} и y_{i+1} .

Например, при $k = 3$ в (8.54), т.е. при использовании первых, вторых и третьих разностей, из (8.55) и (8.58) получаем для задачи (8.49)–(8.50) следующий **явный четырехшаговый метод Адамса четвертого порядка** (в ординатном виде):

$$\begin{cases} y'_{i+1} = y'_i + \frac{h}{24}(55f_i - 59f_{i-1} + 37f_{i-2} - 9f_{i-3}), \\ y_{i+1} = y_i + hy'_i + \frac{h^2}{360}(323f_i - 264f_{i-1} + 159f_{i-2} - 38f_{i-3}). \end{cases} \quad (8.59)$$

Главная часть локальной ошибки первой из формул (8.59) составляет величину $\frac{251}{720}h\Delta^4 f_{i-4} = O(h^5)$, а второй — величину

$$\frac{3}{32}h^2\Delta^4 f_{i-4} = O(h^6).$$

Нетрудно вывести и неявные методы Адамса для задачи (8.49)–(8.50), подменяя функцию f в равенствах (8.53) и (8.56) интерполяционным многочленом вида (8.54) с увеличенными на единицу индексами, т.е. многочленом $P_k(x_{i+1} + qh)$. Совершенно очевидно, что для вычисления y'_{i+1} при этом будет использована выведенная ранее интерполяционная формула Адамса–Моултона (8.17), где вместо y_i , y_{i+1} следует записать соответственно y'_i , y'_{i+1} . Вывод неявной формулы для вычисления y_{i+1} предоставим читателю. На базе явных и неявных методов можно устраивать предиктор-корректорные алгоритмы с пошаговым контролем точности.

Рассмотрим, наконец, один частный вид уравнений (8.49), для которых вычисления по методу Адамса можно сделать более лаконичными. Именно, будем строить метод решения задачи Коши для уравнения

$$y'' = f(x, y), \quad (8.60)$$

правая часть которого не содержит производной. Обратив внимание на то, что в таком случае при вычислении значений f_j значения производной не требуются, вместо двух формул, реализующих явный метод Адамса (8.55), (8.58), можно попытаться ограничиться одной формулой (8.58), если удастся избавиться в

^{*} Если последними учитываются k -е разности, то в скобках формул (8.55) и (8.58) берется $k + 1$ слагаемых, что как раз соответствует шаговости и порядку метода.

ней от слагаемого hy'_i . С этой целью выведем вспомогательную формулу типа (8.58), также содержащую слагаемое hy'_i , что позволит исключить его алгебраическим сложением формул.

Проинтегрируем равенство (8.53), в котором вместо функции $f(x, y, y')$ будем подразумевать функцию $f(x, y)$ — правую часть уравнения (8.60), на промежутке $[x_{i-1}, x_i]$. Имеем

$$y(x_i) = y(x_{i-1}) + hy'(x_i) + \int_{x_{i-1}}^{x_i} \int_x^t f(t, y(t)) dt dx. \quad (8.61)$$

Подставляя сюда y_{i-1} вместо $y(x_{i-1})$, y'_i вместо $y'(x_i)$ и тот же многочлен Ньютона (8.54) вместо подынтегральной функции f , получим формулу, аналогичную (8.57):

$$y_i = y_{i-1} + hy'_i + \int_{x_{i-1}}^{x_i} \int_x^t P_k(t) dt dx. \quad (8.62)$$

С теми же заменами переменных $x = x_i + ph$ и $t = x_i + qh$ производим двойное интегрирование в (8.62) (см. выкладки при выводе формулы (8.58)):

$$\begin{aligned} \int_{x_{i-1}}^{x_i} \int_x^t P_k(t) dt dx &= h^2 \int_{-1}^0 dp \int_0^p P_k(x_i + qh) dq = \\ &= h^2 \left(-\frac{1}{2} f_i + \frac{1}{6} \Delta f_{i-1} + \frac{1}{24} \Delta^2 f_{i-2} + \frac{1}{45} \Delta^3 f_{i-3} + \frac{7}{480} \Delta^4 f_{i-4} + \dots \right). \end{aligned}$$

Будем теперь рассматривать совместно получающуюся подстановкой в (8.62) этого выражения двойного интеграла формулу

$$\begin{aligned} y_i = y_{i-1} + h y'_i + \frac{h^2}{2} \left(-f_i + \frac{1}{3} \Delta f_{i-1} + \frac{1}{12} \Delta^2 f_{i-2} + \right. \\ \left. + \frac{2}{45} \Delta^3 f_{i-3} + \frac{7}{240} \Delta^4 f_{i-4} + \dots \right) \quad (8.63) \end{aligned}$$

и формулу (8.58), где, как и в (8.63), считаем $f_j := f(x_j, y_j)$. Вычитая из равенства (8.58) равенство (8.63), приходим к формуле

$$\begin{aligned} y_{i+1} = 2y_i - y_{i-1} + h^2 \left(f_i + \frac{1}{12} \Delta^2 f_{i-2} + \right. \\ \left. + \frac{1}{12} \Delta^3 f_{i-3} + \frac{19}{240} \Delta^4 f_{i-4} + \dots \right), \quad (8.64) \end{aligned}$$

которая называется **формулой Штёрмера***^{*)} или **Адамса–Штёрмера**. Как видим, **метод Штёрмера** решения начальной задачи (8.60), (8.50) определяется одной явной формулой (8.64) и является более простым и экономичным по сравнению с базовым для него методом Адамса.

Так же просто выводится и неявная формула Штёрмера, которая может быть привлечена к организации вычислений по схеме «предсказание-уточнение».

Замечание 8.3. Многошаговые методы прогноза и коррекции высоких порядков позволяют строить алгоритмы с автоматическим выбором шага, отличающиеся априорно высокой эффективностью. Однако практика использования таких простых алгоритмов показывает, что их реальная эффективность значительно ниже предполагаемой из-за частых изменений величины шага. Поэтому непосредственно методы Адамса прогноза и коррекции на промежутках значительной протяженности целесообразно применять в тех случаях, когда вычисления можно проводить с некоторым наперед известным шагом. В противном случае применяют более эффективные, создаваемые на той же многошаговой основе, специальные алгоритмы Гира [4, 91].

УПРАЖНЕНИЯ

8.1. Запишите явные и неявные формулы Адамса пятого порядка точности (в ординатном виде).

8.2. Получите приближенные критерии пошагового контроля точности в предиктор-корректорных методах Адамса:

- а) первого порядка;
- б) второго порядка;
- в) третьего порядка.

8.3. Интегро-интерполяционным способом выведите предиктор-корректорные формулы третьего порядка точности, используя интегральное равенство типа (8.3) на промежутке $[x_{i-2}, x_{i+1}]$.

Сравните пошаговую точность полученного метода с аналогичным показателем предиктор-корректорного метода Адамса третьего порядка (см. упр. 8.2 в)).

8.4. Выведите вторую формулу Милна и главный член ее локальной ошибки, интегрируя уравнение $y' = f(x, y)$ на промежутке

^{*)} Штёрмер (Стёрмер) Фредрик Карл Мюлерц (1874–1957) — норвежский геофизик и математик.

$[x_{i-1}, x_{i+1}]$ и приближая подынтегральную функцию $f(x, y(x))$ первым многочленом Ньютона третьей степени с базовым узлом x_{i-1} .

8.5. Проверьте выполнение условий согласованности параметров в методах Адамса, Милна и в методе, выведенном в упр.8.3.

8.6. Исходя из общего вида линейных многошаговых методов и понятия алгебраического порядка точности, опишите различные семейства трехшаговых методов разных порядков.

8.7. Для задачи, рассматривавшейся в примере 7.1 § 7.3, выполните один полный шаг метода Коуэлла (8.47), считая $h = 0.1$. Окончательный результат $y_2 \approx y(0.2)$ и все промежуточные результаты последовательных сближений сравните с точными значениями решения в соответствующих точках.

8.8. А) Убедитесь в справедливости формулы Коуэлла (8.47).

Б) Постройте метод не ниже третьего порядка точности на основании интерполяционной формулы Стирлинга (1.32).

8.9. Запишите расчетные формулы получения приближенных значений решения задачи Коши для двумерной системы ОДУ предиктор-корректорным методом Адамса второго порядка. Выполните по этим формулам расчеты для задачи, поставленной в примере 8.1 § 8.5.

8.10. А) Выведите неявную конечноразностную формулу Адамса для численного решения задачи

$$y'' = f(x, y, y'), \quad y(x_0) = y_0, \quad y'(x_0) = y'_0.$$

Б) Запишите четырехшаговый предиктор-корректорный метод Адамса для этой задачи и выведите приближенное правило пошагового контроля точности.

8.11. Запишите частные формулы Штёрмера в ординатном виде:

- а) двухшаговые;
- б) трехшаговые;
- в) четырехшаговые.

8.12. А) Выведите неявную конечноразностную формулу Штёрмера, рассмотрите ее частные случаи.

Б) Запишите четырехшаговый предиктор-корректорный метод Адамса–Штёрмера и получите формулу для пошагового контроля точности этого метода.

ГЛАВА 9 || О ПРОБЛЕМАХ ЧИСЛЕННОЙ || УСТОЙЧИВОСТИ

Описывается общий принцип построения методов решения задач численного анализа и определяются связанные с этим понятия аппроксимации, устойчивости, сходимости. Показывается реализация этого принципа при аппроксимации разностными уравнениями начальных задач для обыкновенных дифференциальных уравнений. Приводятся краткие сведения о решениях линейных разностных уравнений с постоянными коэффициентами, которые затем используются при изучении поведения решений и ошибок некоторых простейших схем. Вводится определение устойчивости по Дальквисту, дающее возможность легко выявлять неустойчивые многошаговые методы. С помощью примеров формируется представление о жестких начальных задачах, дается определение жесткой системы. Через понятие области устойчивости определяется A -устойчивость и $A(\alpha)$ -устойчивость разностных методов, обеспечивающие приемлемость этих методов для решения жестких задач (в частности, выделяются чисто неявные методы дифференцирования назад, обладающие таким свойством).

§ 9.1. ОБЩАЯ СХЕМА РЕШЕНИЯ ЗАДАЧ ЧИСЛЕННОГО АНАЛИЗА. АППРОКСИМАЦИЯ, УСТОЙЧИВОСТЬ, СХОДИМОСТЬ

Большинство задач численного анализа, будь то начальные или граничные задачи для дифференциальных уравнений, интегральные уравнения и т.п., достаточно естественно можно записать в виде уравнения

$$F(y) = z, \quad (9.1)$$

где $F: Y \rightarrow Z$ — линейный или нелинейный оператор (или функционал), переводящий элементы метрического пространства Y в метрическое пространство Z . Суть приближенных методов решения таких задач, как ясно видно из всего предшествующего материала, заключается в том, что уравнение (9.1) заменяется близким ему, в некотором смысле более простым (обычно конечномерным) уравнением

$$F_n(y_n) = z_n. \quad (9.2)$$

Это уравнение определяется оператором $F_n: Y_n \rightarrow Z_n$, соответствующим данному оператору F и действующим из метрического

пространства Y_n в метрическое пространство Z_n . При этом элементы $y_n \in Y_n$ и $z_n \in Z_n$ рассматриваются как образы элементов $y \in Y$ и $z \in Z$ соответствующих исходных пространств, и такая связь задается некоторыми **операторами сноса** $\varphi_n: Y \rightarrow Y_n$ и $\psi_n: Z \rightarrow Z_n$, т.е. равенствами

$$y_n = \varphi_n(y), \quad z_n = \psi_n(z).$$

Обратное соответствие между пространствами Y_n и Y , а точнее, между Y_n и некоторым подпространством пространства Y , устанавливается с помощью **оператора восполнения** φ_n^{-1} :

$$y = \varphi_n^{-1}(y_n). \quad (9.3)$$

Схематично связь между четырьмя фигурирующими здесь пространствами показана на рис.9.1, а).

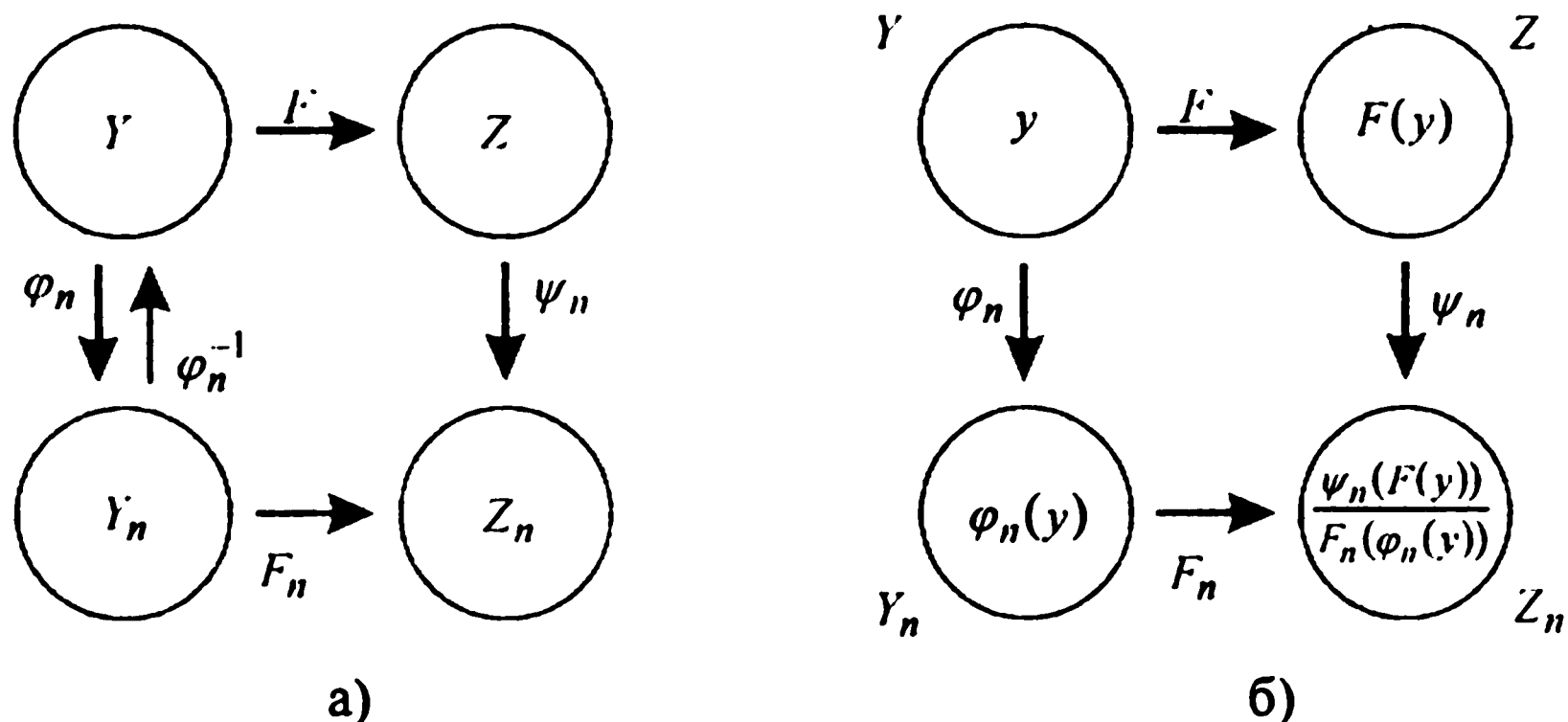


Рис. 9.1. К связи между задачами (9.1) и (9.2)

Чтобы понять, как можно описать близость задач (9.1) и (9.2), проследим связь между пространствами Y , Z , Z_n и Y_n на уровне элементов (рис.9.1, б).

Зафиксируем некоторый элемент $y \in Y$. Его образом в пространстве Z , благодаря данному оператору F , будет элемент $z = F(y)$, а в пространстве Z_n с помощью оператора сноса ψ_n получаем элемент $\psi_n(F(y))$. С другой стороны, тому же элементу y оператор сноса φ_n ставит в соответствие элемент $y_n = \varphi_n(y)$ пространства Y_n , а ему, в свою очередь, новый оператор F_n сопоставляет элемент $F_n(\varphi_n(y))$ пространства Z_n . Так как элементы $\tilde{z}_n := \psi_n(F(y))$ и $\hat{z}_n := F_n(\varphi_n(y))$ служат образами

одного и того же элемента y из Y в одном и том же пространстве Z_n , то по близости между ними можно судить о том, насколько близки операторы F_n и F .

Определение 9.1. Говорят, что уравнение

$$F_n(y_n) = \psi_n(z) \quad (9.4)$$

аппроксимирует уравнение (9.1) (оператор F_n аппроксимирует оператор F), если для любых y из $D(F) \subseteq Y$ мера аппроксимации

$$\rho_{Z_n}(F_n(\varphi_n(y)), \psi_n(F(y))) \quad (9.5)$$

стремится к нулю при $n \rightarrow \infty$.

(Здесь $\rho_{Z_n}(\cdot, \cdot)$ обозначает метрику, т.е. расстояние между указанными в скобках элементами пространства Z_n).

Чтобы иметь возможность грубо сравнивать качество различных моделей типа (9.4) задачи (9.1), часто используют понятие **порядка аппроксимации**, связывая стремление к нулю меры аппроксимации (9.5) с порядком убывания какой-либо зависящей от n малой величины (**шага аппроксимации**).

Предположим, что решения $y^* \in Y$ и $y_n^* \in Y_n$ уравнений соответственно (9.1) и (9.4) существуют и единственны. Поскольку решение задачи (9.1) ищется в пространстве Y , ее **приближенным решением** считается получаемый с помощью оператора восполнения (9.3) элемент

$$y^{(n)} = \varphi_n^{-1}(y_n^*),$$

решение же y_n^* задачи (9.4) называется **каркасом приближенного решения**^{*)}. Наверное, главным вопросом любой теории приближенных методов решения задач вида (9.1) является вопрос о том, можно ли приближенным решением $y^{(n)}$ сколь угодно хорошо отразить поведение точного решения y^* , иначе, вопрос о сходимости $y^{(n)}$ к y^* .

Определение 9.2. Говорят, что имеет место **сходимость приближенных решений** $y^{(n)}$ к точному решению

^{*)} Типичная ситуация: вместо точного решения — функции — находят таблицу ее приближенных значений (каркас), а затем производят восполнение, например, посредством сплайн-интерполирования; естественно, что получаемое при этом приближенное решение принадлежит пространству более узкому, чем исходное пространство решений.

y^* уравнения (9.1), если $\rho_Y(y^*, y^{(n)}) \xrightarrow{n \rightarrow \infty} 0$ (т.е. если расстояние между $y^{(n)}$ и y^* по метрике пространства Y может быть сделано сколь угодно малым).

Очевидно, имеет смысл рассматривать также *сходимость каркасов приближенных решений*, понимая под этим выполнение условия

$$\rho_{Y_n}(y_n^*, \varphi_n(y^*)) \xrightarrow{n \rightarrow \infty} 0.$$

Подобно порядку аппроксимации, вводится понятие порядка сходимости приближенных решений и/или их каркасов.

Наличие фактических оценок величин $\rho_{Y_n}(y_n^*, \varphi_n(y^*))$ позволяет не только делать выводы о сходимости приближенных решений и/или их каркасов, но и указывать погрешности получаемых приближений к решению.

Вопрос о сходимости приближенных решений $y^{(n)}$ к y^* тесно связан с тем, можем ли мы надежно получать каркас решения y_n^* , решая упрощенную задачу (9.4) (в предположении, что последняя аппроксимирует данную задачу (9.1) в смысле определения 9.1)? Дело в том, что эта упрощенная задача (9.4), вообще говоря, другая и тоже решается приближенно; улучшение качества аппроксимации, т.е. уменьшение ее меры (9.5), влечет увеличение размерности n решаемой задачи (9.4), а значит, рост объема вычислений, что, в свою очередь, может привести к катастрофическому нарастанию вычислительных погрешностей.

Определение 9.3. *Вычислительный процесс называется устойчивым, если малые погрешности исходных данных вызывают малые погрешности результата (рост погрешностей ограничен).*

В определении 9.3 заложено скорее понятие, чем строгое определение численной (иначе, вычислительной) устойчивости. Имеется ряд более конкретных определений численной устойчивости применительно к более конкретно поставленным задачам приближенных вычислений (см., например, далее определение 9.4 в § 9.5, определения 9.8 и 9.9 в § 9.7).

Одна из теорем численного анализа гласит: «Аппроксимация плюс устойчивость влечет сходимость». Ее доказательство (и более корректные формулировки) можно найти в книгах [8, 34, 48, 55, 79, 81 и др.].

§ 9.2. ПРОСТЕЙШИЕ РАЗНОСТНЫЕ АППРОКСИМАЦИИ ЗАДАЧИ КОШИ. ГЛОБАЛЬНАЯ ПОГРЕШНОСТЬ МЕТОДА ЭЙЛЕРА

Вернемся к изучению численных процессов решения начальной задачи

$$\begin{aligned} y' &= f(x, y), \quad x \in [x_0, b] \\ y(x_0) &= y_0. \end{aligned} \tag{9.6}$$

Положив $h = \frac{b - x_0}{n}$, введем на отрезке $[x_0, b]$ равномерную сетку^{*})

$$\omega_h := \{ x_i \mid x_i = x_0 + ih \ (i = 0, 1, \dots, n) \}. \tag{9.7}$$

Функции, определенные во всех узлах x_i сетки ω_h , называют **сеточными функциями**. Например, если некоторая функция $y = \varphi(x)$ определена в узлах сетки ω_h , то сеточной функцией следует считать функцию $y_i = \varphi(x_i)$ дискретного аргумента $i = 0, 1, \dots, n$, т.е. $(n+1)$ -мерный вектор $(\varphi(x_0), \varphi(x_1), \dots, \varphi(x_n))$. В свете сказанного в предыдущем параграфе бесконечномерная дифференциальная задача (9.6), состоящая в нахождении удовлетворяющей ей функции — решения $y = y(x)$, сводится к конечномерной задаче вычисления $(n+1)$ -мерного вектора $(y(x_0), y(x_1), \dots, y(x_n))$ — соответствующей этому решению $y(x)$ сеточной функции $y(x_i)$ на ω_h , другими словами, каркаса решения $y_0 \approx y(x_0), y_1 \approx y(x_1), \dots, y_n \approx y(x_n)$. Одним из основных приемов такого сведения является уже использовавшийся в § 7.2 разностный подход, т.е. рассмотрение уравнения (9.6) в узлах сетки ω_h и аппроксимация в них производной y' через соседние значения сеточной функции, соответствующей исходному решению $y = y(x)$ (иначе, замена дифференциального оператора разностным).

Знание порядков используемых формул аппроксимации производной и их остаточных членов позволяет получать представление о порядке аппроксимации данной бесконечномерной задачи конечномерной, выводить оценки погрешностей приближенных решений (на сетке) и изучать устойчивость и сходимость каркасов решений. В связи с таким подходом к построению конечномерных моделей дифференциальных уравнений, соответствующие численные методы их решения называют **разностными уравнениями** или **разностными схемами**.

^{*}) Естественно, сетки бывают и неравномерными.

Положим в (9.6) $x = x_i$ и к левой части полученного равенства

$$y'(x_i) = f(x_i, y(x_i))$$

применим простейшие аппроксимации первого (6.15), (6.14) и второго (6.19), (6.26), (6.27) порядков. Имеем:

$$\begin{aligned} \frac{y(x_{i+1}) - y(x_i)}{h} &= f(x_i, y(x_i)) + O(h); \\ \frac{y(x_i) - y(x_{i-1}))}{h} &= f(x_i, y(x_i)) + O(h); \end{aligned} \quad (9.8)$$

$$\frac{y(x_{i+1}) - y(x_{i-1}))}{2h} = f(x_i, y(x_i)) + O(h^2); \quad (9.9)$$

$$\frac{-3y(x_i) + 4y(x_{i+1}) - y(x_{i+2}))}{2h} = f(x_i, y(x_i)) + O(h^2);$$

$$\frac{y(x_{i-2}) - 4y(x_{i-1}) + 3y(x_i))}{2h} = f(x_i, y(x_i)) + O(h^2). \quad (9.10)$$

Отбрасывание в этих равенствах последнего слагаемого, характеризующего порядок аппроксимации, и замена в них точных значений $y(x_j)$ решения $y(x)$ в j -х узлах сетки ω_h приближенными значениями y_j приводит к следующим разностным схемам:

$$y_{i+1} = y_i + hf(x_i, y_i), \quad i = 0, 1, \dots, n-1; \quad (9.11)$$

$$y_{i+1} = y_i + hf(x_{i+1}, y_{i+1}), \quad i = 0, 1, \dots, n-1; \quad (9.12)$$

$$y_{i+1} = y_{i-1} + 2hf(x_i, y_i), \quad i = 1, 2, \dots, n-1; \quad (9.13)$$

$$y_{i+1} = 4y_i - 3y_{i-1} - 2hf(x_{i-1}, y_{i-1}), \quad i = 1, 2, \dots, n-1; \quad (9.14)$$

$$y_{i+1} = \frac{4}{3}y_i - \frac{1}{3}y_{i-1} + \frac{2h}{3}f(x_{i+1}, y_{i+1}), \quad i = 1, 2, \dots, n-1. \quad (9.15)$$

Первые три из схем (9.11)–(9.15) определяют хорошо знакомые методы Эйлера: явный (7.8), неявный (7.14) и уточненный (7.19). Последние две схемы второго порядка представляют собой явный и неявный двухшаговые методы второго порядка, внимание которым будет уделено позже.

До сих пор нами изучались лишь локальные ошибки методов, т.е. ошибки, возникающие на одном текущем шаге в предположении, что исходным материалом для получения результата

этого шага служат точные значения. Попробуем теперь хотя бы в простейших случаях изучить поведение ошибок в их взаимосвязи на соседних шагах и получить представление о процессе накопления методической погрешности к n -му шагу. Знание того, как изменяются погрешности от шага к шагу, позволяет делать выводы о численной устойчивости разностного метода и о сходимости поставляемых им каркасов приближенных решений, получать оценки глобальной погрешности.

Обозначим

$$\delta_i := y(x_i) - y_i \quad (9.16)$$

— разность между значением точного решения $y = y(x)$ задачи (9.6) в i -м узле сетки (9.7) и соответствующей компонентой каркаса решения, получаемого тем или иным разностным методом.

Предположим, что правая часть и решение данной задачи (9.6) обладает достаточной гладкостью и выполняются условия

$$\exists C_1, C_2 > 0 : |f'_y(x, y)| \leq C_1, \quad |y''(x)| \leq C_2 \quad \forall x \in [x_0, b]. \quad (9.17)$$

В этом предположении оценим глобальную погрешность метода Эйлера.

Вычитая из линейризованного по формуле Тейлора выражения решения $y = y(x)$ в точке x_{i+1} получаемое явным методом Эйлера (9.11) значение y_{i+1} , в соответствии с обозначением (9.16) имеем:

$$\begin{aligned} \delta_{i+1} &= y(x_{i+1}) - y_{i+1} = \\ &= y(x_i) + y'(x_i)h + \frac{1}{2}y''(\Theta_i)h^2 - y_i - hf(x_i, y_i) = \\ &= \delta_i + h[f(x_i, y(x_i)) - f(x_i, y_i)] + \frac{1}{2}y''(\Theta_i)h^2. \end{aligned}$$

Применив к разности функций в последнем выражении формулу Лагранжа по второму аргументу, получаем

$$\delta_{i+1} = \delta_i + hf'_y(x_i, v_i)\delta_i + \frac{1}{2}y''(\Theta_i)h^2. \quad (9.18)$$

Таким образом, связь между ошибками в $(i+1)$ -м и i -м узлах описывается разностным уравнением

$$\delta_{i+1} = A_i\delta_i + B_i, \quad i = 0, 1, \dots, n-1, \quad (9.19)$$

где

$$A_i := 1 + hf'_y(x_i, v_i), \quad B_i := \frac{1}{2}y''(\Theta_i)h^2 \quad (9.20)$$

(Θ_i и v_i — некоторые точки из области задания и области значений решения $y(x)$ соответственно). Анализировать это уравнение будем позже, а здесь произведем оценивание абсолютной погрешности, используя условия (9.17). Благодаря им из (9.18) получаем рекуррентное неравенство

$$|\delta_{i+1}| \leq A|\delta_i| + B, \quad i = 0, 1, \dots, n-1,$$

где

$$A := 1 + C_1 h, \quad B := \frac{1}{2} C_2 h^2. \quad (9.21)$$

Итерирование этого неравенства дает

$$\begin{aligned} |\delta_{i+1}| &\leq A(A|\delta_{i-1}| + B) + B = A^2|\delta_{i-1}| + (A+1)B \leq \\ &\leq A^2(A|\delta_{i-2}| + B) + (A+1)B = A^3|\delta_{i-2}| + (A^2 + A + 1)B \leq \dots \\ &\dots \leq A^{i+1}|\delta_0| + (A^i + A^{i-1} + \dots + A + 1)B = \frac{A^{i+1} - 1}{A - 1} B, \end{aligned}$$

поскольку $\delta_0 = y(x_0) - y_0 = 0$, согласно начальному условию. При $i = n-1$ в соответствии с обозначением (9.21) отсюда получаем оценку глобальной погрешности метода Эйлера

$$|\delta_n| \leq \frac{(1 + C_1 h)^n - 1}{C_1 h} \cdot \frac{1}{2} C_2 h^2 = \frac{C_2 h}{2C_1} [(1 + C_1 h)^n - 1].$$

Чтобы проще было судить о порядке глобальной погрешности, применим в правой части ее оценки формулу Ньютона n -й степени бинома. В результате имеем

$$|y(b) - y_n| \leq \frac{C_2 h}{2C_1} (nC_1 h + o(h)) = \frac{1}{2} C_2 (b - x_0) h + o(h^2) = O(h).$$

Как видим, глобальная погрешность метода Эйлера имеет первый порядок относительно шага h и совпадает по порядку с погрешностью аппроксимации дифференциальной начальной задачи (9.6) дискретной задачей (9.11) с начальным значением y_0 .

Если при выводе разностных уравнений типа (9.19) для ошибок δ_i ограничиваться указанием только порядка свободных членов, то такие уравнения проще получать сравнением исследуемых разностных схем с приведшими к ним результатами аппроксимаций дифференциального оператора разностным. Продемонстрируем это на неявном (9.12) и уточненном (9.13) методах Эйлера.

Перепишем равенство (9.8) в виде

$$y(x_{i+1}) = y(x_i) + hf(x_{i+1}, y(x_{i+1})) + O(h^2)$$

и вычтем из него (9.12). Используя обозначение (9.16) и формулу конечных приращений Лагранжа, имеем

$$\begin{aligned} \delta_{i+1} &= \delta_i + h[f(x_{i+1}, y(x_{i+1})) - f(x_{i+1}, y_i)] + O(h^2) = \\ &= \delta_i + hf'_y(x_{i+1}, v_{i+1})\delta_{i+1} + O(h^2), \end{aligned}$$

откуда после разрешения полученного равенства относительно δ_{i+1} приходим к тому же разностному уравнению (9.19), в котором

$$A_i := \frac{1}{1 - hf'_y(x_{i+1}, v_{i+1})}, \quad B_i := \frac{O(h^2)}{1 - hf'_y(x_{i+1}, v_{i+1})}. \quad (9.22)$$

Аналогично, сравнение равенства (9.13), определяющего уточненный метод Эйлера, с эквивалентным (9.9) равенством

$$y(x_{i+1}) = y(x_{i-1}) + 2hf(x_i, y(x_i)) + O(h^3)$$

дает

$$\begin{aligned} \delta_{i+1} &= \delta_{i-1} + 2h[f(x_i, y(x_i)) - f(x_i, y_i)] + O(h^3) = \\ &= \delta_{i-1} + 2hf'_y(x_i, v_i)\delta_i + O(h^3). \end{aligned}$$

Следовательно, ошибка двухшагового метода (9.13), основанного на симметричной формуле второго порядка аппроксимации производной, удовлетворяет трехточечному рекуррентному соотношению

$$\delta_{i+1} = A_i\delta_i + \delta_{i-1} + O(h^3), \quad (9.23)$$

где $A_i := 2hf'_y(x_i, v_i)$.

§ 9.3. КРАТКИЕ СВЕДЕНИЯ О РЕШЕНИЯХ ЛИНЕЙНЫХ РАЗНОСТНЫХ УРАВНЕНИЙ С ПОСТОЯННЫМИ КОЭФФИЦИЕНТАМИ

Как видно из предыдущего параграфа и из предыдущей главы, численное решение дифференциального уравнения сводится к решению уравнения разностного, связывающего приближенные значения исходного решения в нескольких соседних узлах сетки, количество которых (минус один) определяет,

сколькешаговым является численный метод. Разностным уравнениям удовлетворяют и ошибки методов в узлах (см., например, (9.19), (9.23)). Поэтому для анализа приближенных решений дифференциальных уравнений и их погрешностей важно иметь представление о решениях разностных уравнений.

Ограничимся некоторыми первичными сведениями о решениях *линейных разностных уравнений m -го порядка с постоянными коэффициентами*, имеющих вид

$$u_{i+1} = a_1 u_i + a_2 u_{i-1} + \dots + a_m u_{i-m+1} + b. \quad (9.24)$$

Изучение таких уравнений проводится аналогично изучению линейных дифференциальных уравнений с постоянными коэффициентами. Так же, как и в дифференциальном случае, рассматривается соответствующее (9.24) *однородное разностное уравнение*

$$U_{i+1} = a_1 U_i + a_2 U_{i-1} + \dots + a_m U_{i-m+1}, \quad (9.25)$$

ищется его *общее решение* U_i , представляющее собой линейную комбинацию m *фундаментальных решений* U_{ij} ($j = 1, 2, \dots, m$), находится какое-либо *частное решение* $\bar{u}_i$ неоднородного уравнения (9.24); тогда общее решение u_i уравнения (9.24) представляется суммой U_i и $\bar{u}_i$.

Подобно тому, как в случае линейного дифференциального уравнения решение соответствующего однородного ищут в виде показательной функции, здесь нетривиальное решение уравнения (9.25) естественно искать в виде

$$U_i = \lambda^i \quad (9.26)$$

с некоторой неизвестной постоянной λ ($\neq 0$). Подставляя (9.26) в (9.25), имеем

$$\lambda^{i+1} = a_1 \lambda^i + a_2 \lambda^{i-1} + \dots + a_m \lambda^{i-m+1},$$

откуда после деления на λ^{i-m+1} приходим к уравнению

$$\lambda^m = a_1 \lambda^{m-1} + a_2 \lambda^{m-2} + \dots + a_{m-1} \lambda + a_m. \quad (9.27)$$

Это уравнение называют *характеристическим* по отношению к уравнению (9.25) (а значит, и к (9.24)), поскольку уравнение (9.25) имеет решения вида (9.26) только в том случае, если λ есть корень алгебраического уравнения (9.27).

Предположим, что все корни $\lambda_1, \lambda_2, \dots, \lambda_m$ характеристического уравнения (9.27) — действительные различные. Тогда выражения

$$U_{i1} = \lambda_1^i, \quad U_{i2} = \lambda_2^i, \quad \dots, \quad U_{im} = \lambda_m^i$$

образуют полную *систему фундаментальных решений*, и общее решение этого уравнения есть

$$U_i = c_1 \lambda_1^i + c_2 \lambda_2^i + \dots + c_m \lambda_m^i, \quad (9.28)$$

где $c_1, c_2, \dots, c_m$ — произвольные постоянные.

Такой же вид (9.28) будет иметь общее решение уравнения (9.25) и в случае, когда среди корней уравнения (9.27) есть комплексные, но нет кратных. Если же некоторое число λ_j является k -кратным корнем характеристического уравнения, то, опять-таки, подобно дифференциальному случаю, ему будет соответствовать k фундаментальных решений [81]:

$$\lambda_j^i, \quad i \lambda_j^i, \quad \dots, \quad i^{k-1} \lambda_j^i.$$

Обращаясь теперь к исходному неоднородному уравнению (9.24), непосредственной проверкой убеждаемся, что выражение

$$\bar{u}_i = \frac{b}{1 - a_1 - a_2 - \dots - a_m} \quad (9.29)$$

можно считать его частным решением (если в (9.29) знаменатель не обращается в нуль). Таким образом, общее решение разностного уравнения (9.24) есть функция целочисленного аргумента i , имеющая вид

$$u_i = c_1 \lambda_1^i + c_2 \lambda_2^i + \dots + c_m \lambda_m^i + \frac{b}{1 - a_1 - a_2 - \dots - a_m}, \quad (9.30)$$

если λ_j ($j = 1, \dots, m$) — простые корни характеристического уравнения (9.27), и содержащая слагаемые вида

$$c_j \lambda_j^i + c_{j+1} i \lambda_j^i + \dots + c_{j+k-1} i^{k-1} \lambda_j^i,$$

соответствующие каждому k -кратному корню λ_j .

Для разностных уравнений, как и для дифференциальных, также можно ставить начальные и краевые задачи, задавая значения u_i при определенных значениях i , что позволяет из общих решений фиксированием произвольных постоянных выделять частные решения, удовлетворяющие конкретной задаче. В этой главе для нас будут представлять интерес *начальные задачи* для $(m+1)$ -точечных разностных уравнений (9.24), которые можно считать некоторыми m -шаговыми аппроксимациями задачи Коши (9.6).

§ 9.4. УСТОЙЧИВОСТЬ И НЕУСТОЙЧИВОСТЬ НЕКОТОРЫХ ПРОСТЕЙШИХ РАЗНОСТНЫХ СХЕМ

Изучение устойчивости численных методов решения начальных задач (9.6) обычно проводят на простом уравнении вида

$$y' = py, \quad (9.31)$$

называемом в данном случае *модельным уравнением*; будем пока считать здесь p вещественным параметром^{*)}. Его общее решение есть

$$y = Ce^{px},$$

и решение соответствующей ему задачи Коши с начальным условием $y(x_0) = y_0$ — функция

$$y = y_0 e^{p(x-x_0)} \quad (9.32)$$

— стремится к нулю, если $p < 0$, и бесконечно растет по абсолютной величине при $p > 0$. Посмотрим, как ведут себя ошибки простейших численных методов, примененных к модельному уравнению (9.31).

Метод Эйлера (9.11) на модельном уравнении (9.31) (с $f(x, y) = py$ и $f'_y(x, y) = p$) допускает ошибку, которая, согласно (9.19), (9.20), удовлетворяет рекуррентному равенству (иначе, разностному уравнению)

$$\delta_{i+1} = (1 + ph)\delta_i + O(h^2), \quad (9.33)$$

где $i = 0, 1, 2, \dots$ и $\delta_0 = 0$. Второе слагаемое в (9.33) связано с погрешностью аппроксимации данного дифференциального уравнения (9.31) разностной схемой

$$y_{i+1} = y_i + hpy_i, \quad (9.34)$$

и его влиянием на численную устойчивость процесса (9.34) можно пренебречь (правда, считать его постоянным, т.е. не зависящим от i , можно лишь условно). Характеристическое уравнение (9.27) для (9.33) при $m = 1$ имеет единственный простой корень

$$\lambda = 1 + ph,$$

определяющий фундаментальное решение

$$\Delta_i = (1 + ph)^i$$

^{*)} Впоследствии (см. § 9.7) условие $p \in \mathbf{R}$ будет заменено условием $p \in \mathbf{C}$, и несколько прояснится смысл рассмотрения моделей именно вида (9.31).

соответствующего однородного уравнения. Частным решением уравнения (9.33), согласно форме (9.29), можно считать

$$\bar{\delta}_i = \frac{O(h^2)}{1 - (1 + ph)} = O(h).$$

Следовательно, представление ошибки, накопленной к $(i + 1)$ -му шагу реализации (9.34) метода Эйлера (9.11), имеет вид

$$\delta_i = C(1 + ph)^i + O(h), \quad (9.35)$$

где $C = \delta_0 - O(h)$, а под δ_0 может пониматься либо нуль при точном стартовом значении $y_0 = y(x_0)$, либо небольшая ошибка приближенного ввода $y_0 \approx y(x_0)$.

Анализируя поведение ошибки (9.35) при $i \rightarrow +\infty$, видим, что ее рост будет ограниченным, если шаг сетки h будет удовлетворять неравенству

$$|1 + ph| \leq 1. \quad (9.36)$$

Ясно, что при положительных p это неравенство не может быть выполнено ни при каких $h > 0$; действительно, если решение растет по абсолютной величине (см. (9.32)), то и погрешность получаемого методом Эйлера приближенного решения неизбежно растет. При отрицательных p неравенство (9.36) равносильно

условию $0 \leq h \leq -\frac{2}{p}$, т.е. допустим любой шаг из промежутка $\left[0, -\frac{2}{p}\right]$.

Таким образом, метод Эйлера (9.11) устойчив на модельном уравнении (9.31), если в этом уравнении $p < 0$ и расчетный шаг метода $h \leq -\frac{2}{p}$. Это ограничение на шаг относит явный метод Эйлера к *условно устойчивым методам*.

Неявный метод Эйлера (9.12) имеет ошибку, которая в соответствии с (9.22) удовлетворяет разностному уравнению

$$\delta_{i+1} = \frac{\delta_i}{1 - ph} + \frac{O(h^2)}{1 - ph}.$$

Решение этого уравнения, согласно (9.30) с учетом того, что

$\lambda = \frac{1}{1-ph}$, можно записать так:

$$\delta_i = C\lambda^i + \frac{O(h^2)}{(1-ph)\left(1 - \frac{1}{1-ph}\right)} = C\left(\frac{1}{1-ph}\right)^i + O(h).$$

Отсюда видно, что для невозрастания ошибки δ_i с ростом i нужно потребовать выполнения неравенства

$$\frac{1}{|1-ph|} \leq 1,$$

что при $ph \neq 1$ равносильно неравенству

$$|1-ph| \geq 1. \quad (9.37)$$

Для отрицательных p это неравенство выполняется при любых $h > 0$, т.е. неявный метод Эйлера *абсолютно устойчив*^{*}. Если же $p > 0$, то в таком случае равносильным (9.37) является неравенство $ph \geq 2$, т.е. рост погрешности будет заведомо ограниченным при условии, что расчетный шаг не слишком мал, а именно, при $h \geq \frac{2}{p}$.

Уточненный метод Эйлера (9.13), приложенный к модельному уравнению (9.31), задается однородным трехточечным разностным уравнением второго порядка

$$y_{i+1} = 2phy_i + y_{i-1}, \quad (9.38)$$

а ошибка δ_i , накапливаемая этим методом к i -му шагу, согласно (9.23), удовлетворяет неоднородному уравнению

$$\delta_{i+1} = 2ph\delta_i + \delta_{i-1} + O(h^3). \quad (9.39)$$

Составив характеристическое уравнение вида (9.27) при $m = 2$

$$\lambda^2 = 2ph\lambda + 1$$

и найдя его корни

$$\lambda_1 = ph + \sqrt{1 + p^2h^2} \quad \text{и} \quad \lambda_2 = ph - \sqrt{1 + p^2h^2},$$

^{*} Заметим, что понятия абсолютной и условной устойчивости численных процессов решения задач Коши обычно вводят применительно к асимптотически устойчивым решениям, т.е. для случая $p < 0$ [81].

видим, что $\lambda_1 > 1$ при $p > 0$, а $\lambda_2 < -1$ при $p < 0$. Значит, соответствующее записи (9.30) общее решение уравнения (9.39)

$$\delta_i = C_1 \lambda_1^i + C_2 \lambda_2^i + O(h^2)$$

с ростом i будет расти по абсолютной величине при любых p и h вне зависимости от начальных условий, определяющих ненулевые постоянные C_1 и C_2 .

Таким образом, *уточненный метод Эйлера* (9.13), обладая более высоким порядком аппроксимации, чем явный или неявный методы Эйлера, является *неустойчивым методом*.

Попытаемся понять природу такой неустойчивости метода (9.13).

Сравнивая уравнение (9.38) для каркаса решения и уравнение (9.39) для его ошибки, приходим к выводу, что они имеют одни и те же фундаментальные решения

$$\lambda_1^i = \left(ph + \sqrt{1 + p^2 h^2} \right)^i \quad \text{и} \quad \lambda_2^i = \left(ph - \sqrt{1 + p^2 h^2} \right)^i.$$

Применив к $\left(1 + (ph)^2 \right)^{\frac{1}{2}}$ биномиальное разложение и увидев в результате такого разложения несколько первых членов ряда для экспоненты, при малых ph имеем:

$$\lambda_1^i = \left[ph + 1 + \frac{1}{2}(ph)^2 + o((ph)^2) \right]^i \approx \left(e^{ph} \right)^i = e^{iph}, \quad (9.40)$$

$$\lambda_2^i = \left[ph - 1 - \frac{1}{2}(ph)^2 + o((ph)^2) \right]^i \approx \left(-e^{-ph} \right)^i = (-1)^i e^{-iph}. \quad (9.41)$$

Если теперь рассмотреть точное решение (9.32) модельного уравнения (9.31) на сетке

$$x_0 = 0, \quad x_i = ih,$$

а точнее, фундаментальное решение $Y(x) = e^{px}$ линейного уравнения (9.31), то оказывается, что сеточное фундаментальное решение описывается равенством

$$Y(x_i) = e^{iph}, \quad (9.42)$$

т.е. совпадает с приближенным представлением одного из фундаментальных решений разностного уравнения (9.38).

Итак, трехточечное разностное уравнение (9.38), являясь уравнением второго порядка, имеет два фундаментальных решения: (9.40) и (9.41), одно из которых является *паразитным*. При p отрицательных, когда точное решение (9.42) убывает, за счет паразитного фундаментального решения (9.41) происходит рост приближенного решения y_i . Если же p положительно, то

ошибка δ_i , как уже выяснилось, растет, но поскольку в этом случае растет и решение, рост ошибки не страшен (паразитное фундаментальное решение (9.41) при этом затухает, что влечет убывание относительной погрешности приближенного решения).

Возвращаясь к общему случаю уравнения $y' = f(x, y)$, отметим, что в роли параметра p в приведенных и в аналогичных им исследованиях численной устойчивости методов, согласно (9.20), (9.22) и т.п., должны фигурировать значения функции $f'_y(x, y)$. Если можно считать, что $f'_y(x, y) \approx \text{const}$, то допустимо использование (кусочно-) постоянных аппроксимаций f'_y и применение фактов теории линейных разностных уравнений с постоянными коэффициентами. В противном случае требуется привлечение более тонких результатов о решениях линейных разностных уравнений с переменными коэффициентами.

§ 9.5. ИССЛЕДОВАНИЕ УСТОЙЧИВОСТИ МНОГОШАГОВЫХ МЕТОДОВ

Будем теперь рассматривать на предмет устойчивости $(2m+1)$ -параметрическое семейство линейных m -шаговых методов вида

$$y_{i+1} = \sum_{j=1}^m \alpha_j y_{i+1-j} + h \sum_{j=0}^m \beta_j f(x_{i+1-j}, y_{i+1-j}) \quad (9.43)$$

(см. (8.35)) в предположении, что выполняются условия согласованности (8.37), (8.39) его коэффициентов α_i ($i = 1, 2, \dots, m$) и β_i ($i = 0, 1, \dots, m$), необходимые и достаточные для обеспечения, как минимум, первого порядка точности метода (9.43). Даже на модельном уравнении (9.31) сложно проанализировать накопление погрешностей в численном процессе (9.43) так, как это делалось для простейших разностных схем. Поэтому к исследованию устойчивости многошаговых методов часто применяют упрощенный подход, предложенный в пятидесятых годах двадцатого века шведским математиком Дальквистом. Суть подхода состоит в том, что нелинейное разностное уравнение (9.43), аппроксимирующее данное дифференциальное уравнение (9.6), в свою очередь, аппроксимируется однородным линейным разностным уравнением с постоянными коэффициентами

$$Y_{i+1} = \sum_{j=1}^m \alpha_j Y_{i+1-j}. \quad (9.44)$$

Оно получается отбрасыванием в уравнении (9.43) второго слагаемого из тех соображений, что сходимость метода (а значит, и поведение каркаса решения, поведение ошибки) изучается при $h \rightarrow 0$; наличие множителя h во втором слагаемом позволяет допустить, что оно играет ограниченную роль, если функция f ограничена.

В соответствии с изложенным в двух предыдущих параграфах, поведение решений Y_i разностного уравнения (9.44) и его аппроксимационные свойства по отношению к данному дифференциальному уравнению тесно связаны с величинами корней характеристического уравнения

$$\lambda^m - \alpha_1 \lambda^{m-1} - \dots - \alpha_{m-1} \lambda - \alpha_m = 0, \quad (9.45)$$

через которые выражаются фундаментальные решения уравнения (9.44) и аналогичного ему уравнения для ошибок.

Определение 9.4. *Метод (9.43) при выполнении условий согласованности (8.37), (8.39) называется устойчивым по Дальквисту, если все корни характеристического уравнения (9.45) по модулю не превосходят единицы и среди корней λ_k таких, что $|\lambda_k| = 1$, нет кратных^{*)}. Если, кроме того, $m-1$ корней уравнения (9.45) по модулю меньше единицы, то метод (9.43) называется строго устойчивым [73] или сильно устойчивым [4, 7].*

Чтобы осмыслить определение устойчивости по Дальквисту, достаточно вспомнить, что приближенное выражение ошибки δ_i метода (9.43) представляется линейной комбинацией фундаментальных решений δ_{ij} ($j = 1, 2, \dots, m$) разностного уравнения

$$\delta_{i+1} = \sum_{j=1}^m \alpha_j \delta_{i+1-j},$$

и чтобы не наблюдалось роста ошибки, требуется ограничить единицей корни характеристического уравнения (9.45), через степени которых выражаются фундаментальные решения δ_{ij} . При этом нельзя допустить наличия кратных корней с модулями, равными единице, ибо в противном случае неизбежен рост ошибки при $i \rightarrow \infty$ из-за наличия в представлении δ_i слагаемых с модулями, пропорциональными $i, i^2, \dots, i^{k-1}$, где k — показатель кратности корня. Строгая же устойчивость означает, что в

^{*)} В [81] это требование называют *условием корней*, в [2, 91] — *корневым условием*, а в [8] — *условием α* .

представлении решения Y_i однородного разностного уравнения (9.44) через фундаментальные решения Y_{ij} ($j = 1, 2, \dots, n$) лишь одно слагаемое должно аппроксимировать решение дифференциального уравнения (9.6), а остальные слагаемые должны стремиться к нулю, т.е. паразитные фундаментальные решения должны быть затухающими.

Поскольку данное определение устойчивости не учитывает второе слагаемое в формуле (9.43), с его помощью можно делать лишь грубую отбраковку неустойчивых методов.

Применим определение 9.4 к нескольким изучавшимся ранее методам.

Явный и неявный методы Эйлера (9.11), (9.12) имеют характеристическое уравнение $\lambda - 1 = 0$ с единственным корнем $\lambda_1 = 1$; следовательно они строго устойчивы по Дальквисту (как, кстати, и любой другой одношаговый метод, рассматриваемый в качестве частного случая m -шаговых методов).

Уточненный метод Эйлера (9.13) является двухшаговым методом с характеристическим уравнением $\lambda^2 - 1 = 0$; наличие двух простых корней $\lambda_{1,2} = \pm 1$ говорит об устойчивости этого метода по Дальквисту, но не строгой устойчивости (ср. с более детальным изучением его устойчивости, точнее сказать, неустойчивости, в предыдущем параграфе).

Метод Адамса (явный или неявный)

$$y_{i+1} = y_i + h \sum_{j=0}^m \beta_j f(x_{i+1-j}, y_{i+1-j})$$

строго устойчив по Дальквисту при любом $m \in \mathbb{N}$, так как его характеристическое уравнение $\lambda^m - \lambda^{m-1} = 0$ имеет один корень, равный единице, а остальные $m - 1$ корней равны нулю.

То же можно сказать и о методе Коуэлла (8.47).

Метод Милна четвертого порядка, как известно, определяется двумя формулами:

прогноза (8.28)

$$\hat{y}_{i+1} = y_{i-3} + \frac{4h}{3}(2f_i - f_{i-1} + 2f_{i-2})$$

и коррекции (8.31)

$$y_{i+1} = y_{i-1} + \frac{h}{3}(\hat{f}_{i+1} + 4f_i + f_{i-1}).$$

Первая из этих формул является четырехшаговой с характеристическим уравнением $\lambda^4 - 1 = 0$, корни которого $\lambda_{1,2} = \pm 1$, $\lambda_{3,4} = \pm i$, а вторая — двухшаговой, для которой соответственно имеем $\lambda^2 - 1 = 0$, $\lambda_{1,2} = \pm 1$. Как видим, та и другая формулы определяют разностный метод, устойчивый по Дальквисту, но не строго устойчивый. Таким образом, метод Милна четвертого порядка, несколько выигрывая у метода прогноза и коррекции (8.22), построенного на основе методов Адамса того же порядка, по точностной характеристике (что отмечалось в конце § 8.3), проигрывает последнему по устойчивости.

Требование устойчивости по Дальквисту должно учитываться при конструировании конкретных методов из семейства (9.43). Это означает, что при подборе параметров в формуле (9.43) следует заботиться не только о том, чтобы она аппроксимировала данное дифференциальное уравнение как можно точнее, но и чтобы соответствующее этой формуле алгебраическое уравнение (9.45) имело ограниченные единицей модули корней и не допускались кратные корни с модулями, равными единице.

Априори $2m+1$ коэффициентов α_j, β_j в методе (9.43) можно подобрать так, чтобы находимые с его помощью приближения y_i аппроксимировали значения решения $y(x_i)$ задачи (9.6) с порядком $2m$, т.е. чтобы этот метод был точен для многочленов степени $2m$ (такие примеры можно найти в § 8.4). Однако доказано, что построенные на этой основе методы наивысшего алгебраического порядка точности являются заведомо неустойчивыми. Этот факт уточняется следующим утверждением [8].

Теорема 9.1. Пусть s — порядок аппроксимации m -шаговым разностным методом (9.43) задачи (9.6). Тогда в каждом из следующих случаев:

- а) метод явный ($\beta_0 = 0$) и $s > m$;
 - б) метод неявный ($\beta_0 \neq 0$), m — нечетное и $s > m + 1$;
 - в) метод неявный ($\beta_0 \neq 0$), m — четное и $s > m + 2$
- среди корней λ_j характеристического уравнения (9.45) найдется корень, по модулю больший единицы.

§ 9.6. ЖЕСТКИЕ УРАВНЕНИЯ И СИСТЕМЫ

Имеются задачи, для которых вопрос об устойчивости или неустойчивости применяемых численных методов стоит наиболее остро и требует большой дифференциации. Речь идет о начальных задачах для дифференциальных уравнений, называемых жесткими. Такие задачи, возникающие в самых разных приклад-

ных областях, в последние десятилетия являются объектом повышенного внимания специалистов по вычислительной математике и служат тем оселком, на котором оттачиваются понятия, формулировки, методы, алгоритмы, программы.

В литературе можно встретить несколько определений жесткости, отличающихся разным уровнем строгости. При первом знакомстве более важно понять, в чем состоит проблема при численном интегрировании дифференциальных уравнений, выделяющая какие-то из них в разряд жестких, и как ведут себя те или иные методы на таких задачах.

Приведем цитату из посвященной жестким уравнениям монографии К. Деккера и Я. Вервера [29]: *«Сущность явления жесткости состоит в том, что решение, которое нужно вычислить, меняется медленно, однако существуют быстро затухающие возмущения. Наличие таких возмущений затрудняет получение медленно меняющегося решения численным способом»*. Попытаемся получить некоторые представления об этом явлении с помощью следующих двух примеров [73].

Пример 9.1. Точным решением задачи Коши

$$\begin{aligned} y' &= -100y + 100, \\ y(0) &= 2 \end{aligned} \tag{9.46}$$

является функция

$$y = 1 + e^{-100x}, \tag{9.47}$$

которая уже при малых положительных x становится близкой к своему предельному значению $y = 1$. Это означает, что для получения численного решения на достаточно большом промежутке $[0, b]$ естественно с мелким шагом построить решение в его переходной фазе (в правой окрестности нуля, т.е. там, где оно быстро меняется), а затем продолжить численный процесс с крупным шагом в области малых изменений решения.

Посмотрим, на что можно рассчитывать, применяя к задаче (9.46) (отличающейся от модельной задачи (9.31) лишь наличием свободного члена) простейшие разностные схемы, устойчивость которых изучалась в § 9.4.

Расчетная формула условно устойчивого метода Эйлера (9.11) для (9.46) имеет вид

$$y_{i+1} = (1 - 100h)y_i + 100h; \quad i = 0, 1, \dots; \quad y_0 = 2. \tag{9.48}$$

В соответствии с условием устойчивости (9.36) расхождение между y_i и $y(x_i)$ не будет расти, если расчетный шаг удовлетворяет неравенству $|1 - 100h| \leq 1$, т.е. если $h \leq 0.02$. Взяв $h = 0.02$, из (9.48) получаем численный процесс

$$y_{i+1} = 2 - y_i; \quad i = 0, 1, \dots; \quad y_0 = 2,$$

порождающий последовательность

$$y(0) = y_0 = 2, \quad y(0.02) \approx y_1 = 0, \quad y(0.04) \approx y_2 = 2, \quad y(0.06) \approx y_3 = 0, \quad \dots$$

Как видим, решение разностного уравнения получается ограниченным, но нельзя считать, что оно удовлетворительно приближает решение (9.47) задачи (9.46) (имеет место «четно-нечетная болтанка» [6]). При $h = 0.01$ (это середина допустимого для h интервала устойчивости $(0, 0.02)$) процесс (9.48) приобретает вид

$$y_0 = 2, \quad y_{i+1} = 1 \quad \forall i \in N_0,$$

показывающий, что первая же вычисляемая точка y_1 ($\approx y(0.01)$) попадает на асимптоту решения (9.47), и последующие вычисления не изменяют значений приближенного решения. Несмотря на устойчивость этого численного процесса и стремящуюся к нулю ошибку при $i \rightarrow \infty$, он совсем не описывает переходный процесс. Существенно более мелкий шаг, например, $h = 0.001$ вполне удовлетворительно характеризует начальное изменение решения, что видно из получающейся при этом последовательности

$$y_0 = 2, \quad y(0.001) \approx y_1 = 1.9, \quad y(0.002) \approx y_2 = 1.81, \dots,$$

но потребует больших вычислительных затрат (и, как следствие, повлечет накопление ошибок округлений) при построении приближенного решения на промежутке, достаточно большом по сравнению с величиной шага h .

Применение здесь абсолютно устойчивого неявного метода Эйлера (9.12) означает проведение вычислений по формуле

$$y_{i+1} = \frac{100h + y_i}{1 + 100h}; \quad i = 0, 1, \dots; \quad y_0 = 2. \quad (9.49)$$

Полагая здесь для удобства счета $h = 0.09$, получаем следующую последовательность приближенных ординат решения (9.47):

$$y_0 = 2, \quad y(0.09) \approx y_1 = 1.1, \quad y(0.18) \approx y_2 = 1.01, \quad y(0.27) \approx y_3 = 1.001, \dots$$

Даже при очень крупном шаге $h = 0.99$ определяемая посредством (9.49) последовательность

$$y_0 = 2, \quad y_1 = 1.01, \quad y_2 = 1.0001, \dots$$

качественно верно определяет поведение решения (9.47) задачи (9.46). Так что для получения приближенного решения данной задачи неявным методом Эйлера с нужной точностью при выборе расчетного шага h достаточно позаботиться лишь об адекватной точности аппроксимации дифференциального уравнения разностным.

Неустойчивая разностная схема (9.13), иначе, уточненный метод Эйлера, в данном примере приобретает вид

$$y_{i+1} = y_{i-1} - 200hy_i + 200h; \quad i = 1, 2, \dots, \quad (9.50)$$

где, кроме известной начальной точки $y_0 = y(0) = 2$, требуется задание еще одной точки y_1 . Положим $h = 0.01$ и, соответственно, за y_1 примем значение

$$y_1 = 1.37 \quad \left(\approx y(0.01) = 1 + \frac{1}{e} \right).$$

В таком случае вычисления по рекуррентной формуле (9.50) продолжают эту последовательность следующим образом:

$$y_2 = 1.26, \quad y_3 = 0.85, \quad y_4 = 1.56, \quad y_5 = -0.27, \quad y_6 = 4.10, \dots$$

Не вызывает сомнений ее непригодность для приближенного описания поведения решения (9.47). Взяв в 100 раз меньший шаг $h = 0.0001$, при котором погрешность аппроксимации дифференциальной задачи (9.46) дискретной задачей (9.50), согласно (9.9), есть $O(10^{-8})$, для тех же начальных элементов последовательности $y_0 = 2$ и $y_1 = 1.37$ имеем продолжение

$$\begin{aligned} y_2 &= 1.993, \quad y_3 = 1.350, \quad y_4 = 1.986, \quad y_5 = 1.330, \\ y_6 &= 1.979, \quad y_7 = 1.311, \quad y_8 = 1.973, \dots \end{aligned}$$

Пульсирующий характер последовательностей с достаточно большими амплитудами пульсаций может быть объяснен поведением паразитной составляющей решения разностного уравнения (9.50), о чем говорилось при анализе причин неустойчивости уточненного метода Эйлера в § 9.4.

Пример 9.2. Уравнение второго порядка

$$y'' + 101y' + 100y = 0 \tag{9.51}$$

с начальными условиями

$$y(0) = 1.01, \quad y'(0) = -2$$

имеет решением функцию

$$y(x) = 0.01e^{-100x} + e^{-x}. \tag{9.52}$$

Первое из слагаемых, составляющих эту функцию, является быстро-затухающим, а второе изменяется сравнительно плавно; проследить за поведением этих двух слагаемых на промежутке $[0, 0.1]$ можно по следующей таблице их приближенных значений:

x	$Y_1 = 0.01e^{-100x}$	$Y_2 = e^{-x}$
0	0.01	1
0.00001	0.009999	0.99999
0.0001	0.0099	0.9999
0.001	0.009	0.999
0.01	0.004	0.99
0.1	0.0000004	0.9

Очевидно, за пределами промежутка $[0, 0.1]$, характеризующего переходную фазу, вклад первого слагаемого Y_1 в решение (9.52) задачи (9.51) ничтожен. Это означает, что при последовательном получении приближенных численных значений решения $y(x)$ промежуток $[0, 0.1]$ должен проходиться с маленьким шагом, а затем, чтобы уменьшить вычислительные затраты и погрешность округлений, расчетный шаг должен быть увеличен. Всякие ли вычислительные методы позволяют произвести укрупнение шага после построения приближенного решения в переходной фазе? Ответ на этот вопрос будет отрицательным.

Действительно, как отмечалось в § 8.6, для построения численного решения задачи (9.51) формально можно применять разные численные методы решения задачи Коши для уравнений первого порядка в векторной форме, сведя ее к системе

$$\begin{cases} y' = z, & y(0) = 1.01, \\ z' = -100y - 101z, & z(0) = -2. \end{cases} \tag{9.53}$$

Использование здесь условно устойчивых явных методов, например, таких как метод Эйлера, накладывает ограничение на величину шага, обусловленное как раз необходимостью устойчивой аппроксимации быстро-затухающей составляющей решения.

Вообще, переход к системе дифференциальных уравнений позволяет взглянуть на проблему жесткости с несколькими позиций.

Пусть n -мерная система

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} \quad (9.54)$$

имеет асимптотически устойчивое решение

$$\mathbf{x}(t) = (x_1(t), \dots, x_n(t))^T.$$

Если его компоненты, т.е. функции $x_1(t), \dots, x_n(t)$, существенно различаются по скорости своего изменения на промежутке $[t_0, T]$, на котором решается задача Коши для (9.54), то применение здесь условно устойчивых численных процессов требует интегрирования с таким малым шагом, какой обеспечивает устойчивое вычисление самой быстро-затухающей компоненты (ведь шаг h — величина скалярная, общая для всех компонент). Следовательно, жесткость системы дифференциальных уравнений зависит от того, насколько сильно разнится поведение компонент вектора-решения при условии его асимптотической устойчивости. В свою очередь, в случае постоянной матрицы $\mathbf{A}$ в системе (9.54) это различие в скорости изменения функций $x_1(t), \dots, x_n(t)$ связано с тем, насколько сильно различаются собственные числа $\lambda_1, \dots, \lambda_n$ матрицы $\mathbf{A}$. Обоснование такой связи можно найти, например, в [4, 81]; возвращаясь же к примеру 9.2 и найдя собственные числа $\lambda_1 = -100$, $\lambda_2 = -1$ матрицы

$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ -100 & -101 \end{pmatrix}$$

системы (9.53), убеждаемся, по крайней мере, в рациональности такого суждения.

Отсюда приходим к следующему определению жесткой системы [81].

Определение 9.5. Система (9.54) с постоянной $n \times n$ -матрицей $\mathbf{A}$ называется **жесткой**, если собственные числа λ_k ($k = 1, 2, \dots, n$) матрицы $\mathbf{A}$ удовлетворяют следующим условиям:

$$1) \operatorname{Re} \lambda_k < 0 \quad \forall k \in \{1, 2, \dots, n\};$$

$$2) \text{ Число жесткости } g := \frac{\max\{|\operatorname{Re} \lambda_k|\}}{\min\{|\operatorname{Re} \lambda_k|\}} \text{ велико.}$$

Нестрогость, заключенная в последнем слове данного определения, сродни той, которая присутствует при введении понятия «плохая обусловленность матрицы» (см., например, [18]); избавиться от нее можно, лишь рассматривая конкретную задачу.

В случае линейной системы с переменными коэффициентами, т.е. при $A = A(t)$ в (9.54), собственные числа λ_k и, соответственно, число жесткости g являются функциями от t , и определение 9.5 жесткой системы может быть переформулировано следующим образом.

Определение 9.6. Система

$$\dot{x} = A(t)x, \quad t > 0$$

называется жесткой на интервале $(0, T)$, если при всех $t \in (0, T)$

$$\operatorname{Re} \lambda_k(t) < 0 \quad \forall k \in \{1, 2, \dots, n\}$$

и число $\sup_{t \in (0, T)} g(t)$ велико.

При определении жесткости нелинейной системы

$$\dot{x} = F(x, t)$$

в основу кладут предыдущее определение, роль матрицы $A(t)$ в котором отводится матрице частных производных $\frac{\partial f_i}{\partial x_j}$ ($i, j = 1, 2, \dots, n$), где f_i и x_i — компоненты вектор-функций F и x соответственно.

Заметим, что требование асимптотической устойчивости точного решения системы дифференциальных уравнений, обеспечиваемое выполнением первого условия приведенных определений жесткости, иногда заменяют более слабым условием асимптотической устойчивости не всех, а только быстро затухающих компонент вектора-решения, что расширяет класс жестких систем. С другой стороны, промежуток, на котором действуют быстро затухающие возмущения, часто не относят к промежутку жесткости [29] (например, в примере 9.1 таковым является промежуток $[0, 0.1]$).

§ 9.7. A - И $A(\alpha)$ -УСТОЙЧИВОСТЬ. ЧИСТО НЕЯВНЫЕ МЕТОДЫ

Как следует из материала предыдущего параграфа, жесткие уравнения предъявляют жесткие требования к устойчивости

численных методов, применяемых для их решения. А именно, при получении асимптотически устойчивого решения жесткой задачи Коши ошибка разностного метода не должна расти при любом шаге, т.е. метод должен быть безусловно устойчивым. Чтобы оформить сказанное более четко, дадим сначала определение области устойчивости [81].

Определение 9.7. *Областью устойчивости разностного метода (9.43) решения начальной задачи (9.6) называется множество всех точек комплексной плоскости, определяемой комплексной переменной $\mu = \rho h$, для которых этот метод, примененный к модельному уравнению (9.31), устойчив, т.е. обеспечивает невозрастание ошибки.*

Теперь сформулируем определение устойчивости метода, ориентированное на применение к жестким задачам.

Определение 9.8. *Разностный метод (9.43) называется A -устойчивым, если его область устойчивости содержит левую полуплоскость комплексной плоскости, определяемой переменной $\mu = \rho h$.*

Обращаясь к простейшим разностным схемам, сразу отметим, что об A -устойчивости явного метода Эйлера (9.11) не может быть и речи. Действительно, согласно проведенным в § 9.4 исследованиям, для невозрастания ошибки этого метода требуется выполнение неравенства (см. (9.36))

$$|\mu + 1| \leq 1,$$

что реализуется в круге радиуса 1 с центром в точке $(-1; 0)$ комплексной плоскости с осями Ou и Ov , где

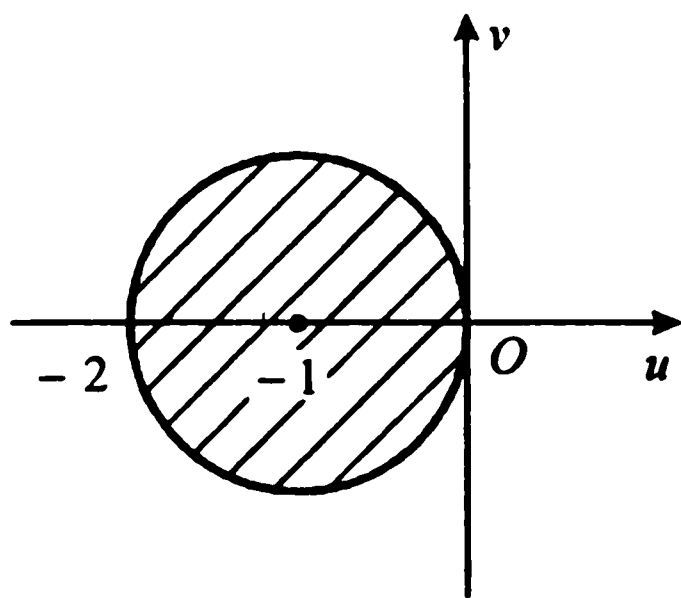
$$u = \operatorname{Re} \mu = \operatorname{Re}(\rho h), \quad v = \operatorname{Im} \mu = \operatorname{Im}(\rho h)$$

(рис. 9.2, а)). Этим неравенством задается его область устойчивости, составляющая лишь малую часть левой полуплоскости плоскости Ouv .

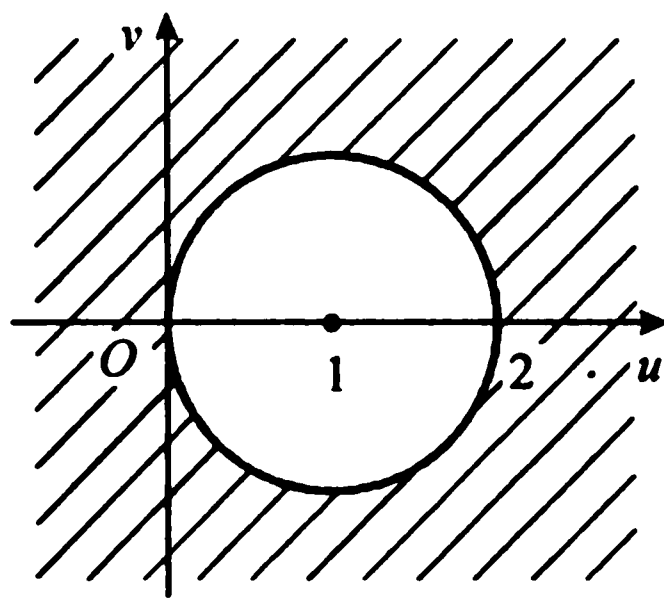
Устойчивость неявного метода Эйлера (9.12), в соответствии с неравенством (9.37), связана с выполнением неравенства

$$|\mu - 1| \geq 1,$$

которое определяет на комплексной плоскости Ouv внешность такого же круга, но круг этот расположен в правой полуплоскости (рис. 9.2, б)). Как видим, область устойчивости неявного метода Эйлера целиком содержит левую полуплоскость; следовательно, *неявный метод Эйлера A -устойчив.*



а)



б)

Рис. 9.2. Области устойчивости явного (а) и неявного (б) методов Эйлера

Рассмотрим на предмет A -устойчивости метод трапеций (7.15) (он же — метод Адамса-Моултона второго порядка (8.19)). На модельном уравнении (9.31) его расчетная формула имеет вид

$$y_{i+1} = y_i + \frac{ph}{2}[y_i + y_{i+1}], \quad i = 0, 1, \dots \quad (9.55)$$

Те или другие источники получения разностной схемы (9.55) позволяют считать, что подстановка в нее точного значения $y(x_i)$ вместо y_i сопровождается ошибкой $O(h^3)$, т.е. имеет место равенство

$$y(x_{i+1}) = y(x_i) + \frac{ph}{2}[y(x_i) + y(x_{i+1})] + O(h^3).$$

Вычитая равенство (9.55) из последнего равенства, приходим к следующему двухточечному разностному уравнению первого порядка относительно ошибки $\delta_i := y(x_i) - y_i$:

$$\left(1 - \frac{ph}{2}\right)\delta_{i+1} = \left(1 + \frac{ph}{2}\right)\delta_i + O(h^3).$$

Отсюда видим, что невозрастание ошибки δ_i гарантируется при выполнении неравенства

$$\left|\frac{2 + ph}{2 - ph}\right| \leq 1.$$

Полагая здесь $\mu = ph = u + iv$, $\mu \neq 2$, имеем

$$|2 + u + iv| \leq |2 - u - iv| \Leftrightarrow \sqrt{(2 + u)^2 + v^2} \leq \sqrt{(2 - u)^2 + v^2}$$

$$\Leftrightarrow u \leq 0 \Leftrightarrow \operatorname{Re} \mu \leq 0.$$

Таким образом, область устойчивости метода трапеций точно совпадает с левой полуплоскостью комплексной плоскости чисел $\mu = \rho h$; следовательно, метод трапеций *A*-устойчив.

Докажем еще *A*-устойчивость неявного двухшагового разностного метода второго порядка (9.15), выведенного в § 9.2.

Аналогично предыдущему, легко получаем разностное уравнение для ошибки каркаса решения модельного уравнения (9.31). Оно имеет вид

$$(3 - 2\rho h)\delta_{i+1} - 4\delta_i + \delta_{i-1} + O(h^3) = 0.$$

Его характеристическое уравнение

$$(3 - 2\mu)\lambda^2 - 4\lambda + 1 = 0 \quad (9.56)$$

имеет корнями числа

$$\lambda_{1,2} = \frac{2 \pm \sqrt{1 + 2\mu}}{3 - 2\mu},$$

непосредственное оценивание которых, точнее, построение области, содержащей только те точки, при которых $|\lambda_{1,2}| \leq 1$, вызывает определенные затруднения. Изберем другой путь [81].

Попытаемся выяснить вид границы области устойчивости, рассматривая равенство (9.56) при таких комплексных λ , модуль которых равен единице. При этом будем пользоваться как тригонометрической, так и показательной формами комплексного числа, т.е. тем, что при $|\lambda| = 1$ по формуле Эйлера

$$\lambda = \cos \varphi + i \sin \varphi = e^{i\varphi}. \quad (9.57)$$

Выражая из (9.56) текущую комплексную переменную μ , с помощью равенства (9.57) получаем:

$$\begin{aligned} \mu &= \frac{1}{2\lambda^2} - \frac{2}{\lambda} + \frac{3}{2} = \frac{1}{2}e^{-2i\varphi} - 2e^{-i\varphi} + \frac{3}{2} = \\ &= \frac{1}{2}\cos 2\varphi - \frac{1}{2}i\sin 2\varphi - 2\cos \varphi + 2i\sin \varphi + \frac{3}{2} = u + iv, \end{aligned}$$

где

$$u := \frac{3}{2} - 2\cos \varphi + \frac{1}{2}\cos 2\varphi = (1 - \cos \varphi)^2 \geq 0, \quad (9.58)$$

$$v := 2\sin \varphi - \frac{1}{2}\sin 2\varphi = \sin \varphi(2 - \cos \varphi). \quad (9.59)$$

Найденные выражения u и v текущей точки μ границы области устойчивости можно рассматривать как ее параметрические уравнения, в которых параметром служит полярный угол φ . Из (9.58) сразу видно, что граница находится в правой полуплоскости, а исследуя определенную в (9.59) функцию v на экстремум, выясняем, что существуют два значения параметра

$\varphi_{1,2} = \arccos \frac{1 \pm \sqrt{3}}{2}$, которым соответствуют экстремальные значения

$v \approx \pm 2$ при $u \approx 1.85$. Найдя еще дополнительно две характерные точки границы $(0;0)$ и $(4;0)$, схематично изображаем замкнутую линию, определяемую параметрическими уравнениями (9.58), (9.59) (рис. 9.3). Эта линия делит всю комплексную плоскость на две области: в одной из них модули корней уравнения (9.56) меньше единицы, в другой — хотя бы один из корней

больше единицы. Подставив в (9.56), например, $\mu = -\frac{1}{2}$, получа-

ем уравнение $4\lambda^2 - 4\lambda + 1 = 0$ с корнями $\lambda_{1,2} = \frac{1}{2} < 1$. Следовательно,

область устойчивости метода (9.15) лежит за пределами построенной замкнутой линии и содержит левую полуплоскость.

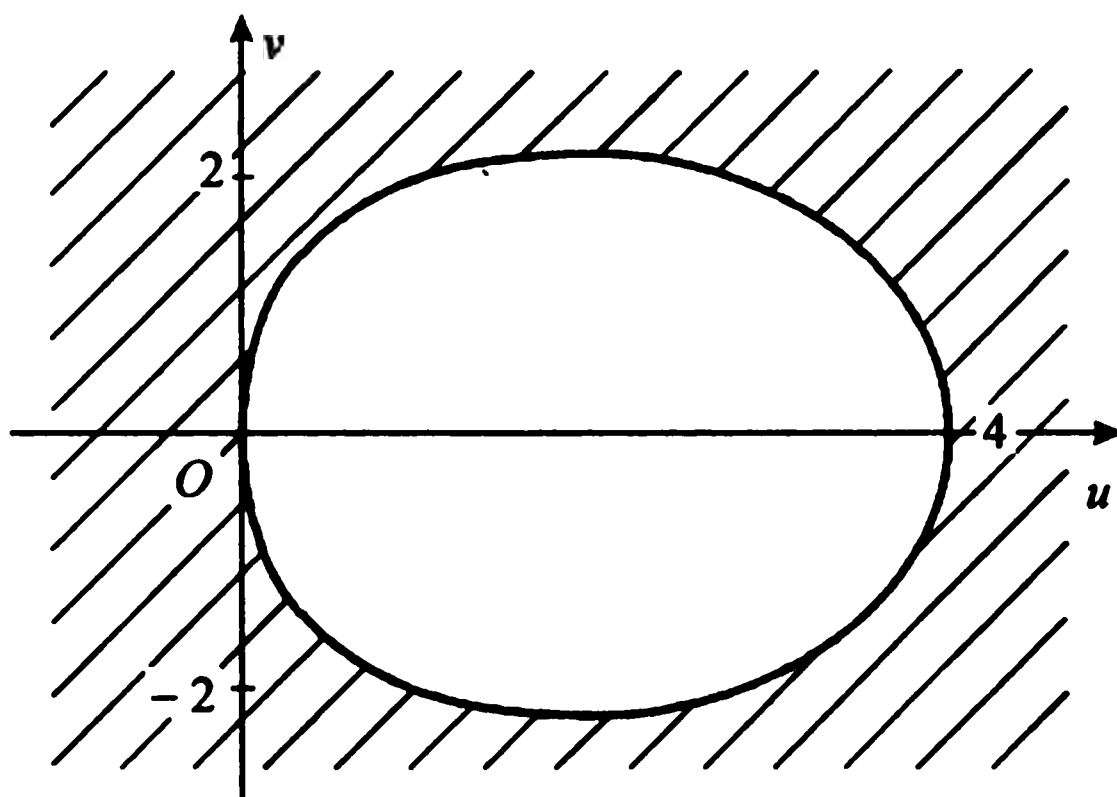


Рис. 9.3. Область устойчивости разностной схемы

$$y_{i+1} = \frac{4}{3}y_i - \frac{1}{3}y_{i-1} + \frac{2h}{3}f(x_{i+1}, y_{i+1})$$

К сожалению, класс A -устойчивых методов весьма узок. Доказано, что среди линейных многошаговых методов (9.43) нет явных A -устойчивых методов, а порядок неявных A -устойчивых методов этого семейства (9.43) не может быть больше двух.

В связи с этим представляет интерес введение более слабых определений устойчивости, которые расширяли бы множество разностных методов, хотя бы ограниченно пригодных для численного решения жестких задач.

Прямым обобщением A -устойчивости является $A(\alpha)$ -устойчивость.

Определение 9.9. Разностный метод (9.43) называется $A(\alpha)$ -устойчивым, если существует угол $\alpha \in \left(0, \frac{\pi}{2}\right]$ такой, что область устойчивости метода (рис. 9.4.) содержит сектор комплексной плоскости переменных $\mu = \rho h$, определяемый неравенством

$$|\arg(-\mu)| < \alpha$$

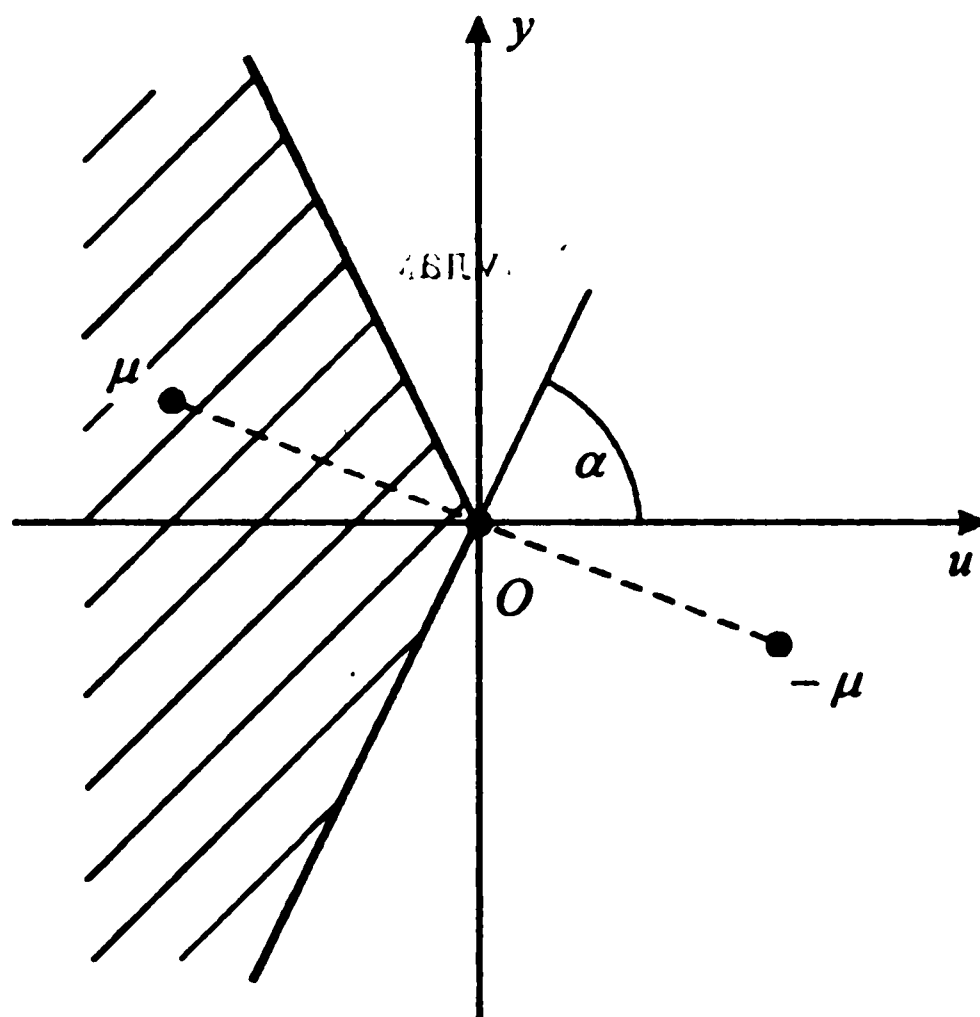


Рис. 9.4. Изображение области (заштрихована), которая должна содержаться в области устойчивости $A(\alpha)$ -устойчивого метода

Очевидно, что A -устойчивость — это $A\left(\frac{\pi}{2}\right)$ -устойчивость.

$A(\alpha)$ -устойчивые методы можно найти в подсемействе семейства m -шаговых разностных методов (9.43), определяемом формулой

$$\sum_{j=0}^m \alpha_j y_{i+1-j} = hf(x_{i+1}, y_{i+1}). \quad (9.60)$$

Счет по формулам вида (9.60) не использует значений производных решения в предыдущих узлах, и поэтому такие методы называют **чисто неявными методами**. Источником получения конкретных $A(\alpha)$ -устойчивых чисто неявных методов вида (9.60) служит применение к аппроксимации производной в равенстве

$$y'(x_{i+1}) = f(x_{i+1}, y_{i+1}) \quad (9.61)$$

формул дифференцирования назад^{*)}, т.е. несимметричных формул, аппроксимирующих $y'(x_{i+1})$ с разным порядком точности с помощью значений y_{i+1}, y_i, y_{i-1} и т.д.

Так, аппроксимация $y'(x_{i+1})$ в (9.61) разностными отношениями

$$\frac{1}{h}(y_{i+1} - y_i) \quad \text{и} \quad \frac{1}{2h}(3y_{i+1} - 4y_i + y_{i-1})$$

приводит к изученным нами $A\left(\frac{\pi}{2}\right)$ -устойчивым чисто неявным методам первого (9.12) и второго (9.15) порядков. Аналогично, аппроксимация $y'(x_{i+1})$ по формулам, например, третьего и четвертого порядков приводит к чисто неявным методам

$$11y_{i+1} - 18y_i + 9y_{i-1} - 2y_{i-2} = 6hf(x_{i+1}, y_{i+1}) \quad (9.62)$$

и

$$25y_{i+1} - 48y_i + 36y_{i-1} - 16y_{i-2} + 3y_{i-3} = 12hf(x_{i+1}, y_{i+1}), \quad (9.63)$$

которые, как доказано, являются $A(\alpha)$ -устойчивыми с $\alpha \approx 1.544$ рад ($\approx 88^\circ$) для метода третьего порядка (9.62) и с $\alpha = \arctg \sqrt{6} \approx 1.278$ рад ($\approx 68^\circ$) для метода четвертого порядка (9.63) [81]. С повышением порядка метода дифференцирования назад значения угла α уменьшаются, и при порядке выше шестого $A(\alpha)$ -устойчивых методов среди семейства многошаговых методов (9.60) не обнаруживается ни для каких $\alpha > 0$ [89].

Заметим, что при применении чисто неявных методов к численному интегрированию жестких задач соответствующие используемому методу разностные уравнения, в общем случае нелинейные относительно очередного значения y_{i+1} , должны решаться достаточно точно. Здесь рекомендуется привлекать

^{*)} В связи с чем и выводимые таким образом методы численного решения задач Коши часто называют **методами дифференцирования назад** [91].

быстросходящиеся итерационные методы типа метода Ньютона [4, 18, 91].

Вернемся к началу этого параграфа и посмотрим, что означают A - и $A(\alpha)$ -устойчивость разностных методов, применяемых именно к жестким системам дифференциальных уравнений.

Если некоторый метод A -устойчив, то он пригоден для формального нахождения приближенного решения модельного уравнения $y' = py$ с любым шагом $h > 0$ при любых комплексных p с отрицательными вещественными частями, а это в свете наблюдений предыдущего параграфа означает, что шаг h может быть подобран так, чтобы система (9.54) решалась устойчиво при любом разбросе собственных значений ее матрицы A , т.е. практически при любой величине числа жесткости g . Если же метод

$A(\alpha)$ -устойчив с $\alpha < \frac{\pi}{2}$, то допустимость этого метода для решения системы (9.54) можно связать с тем, вписываются ли собственные числа λ_k ее матрицы, а точнее, величины $\lambda_k h$, в определяемый данным α сектор комплексной плоскости (см. определение 9.9).

УПРАЖНЕНИЯ

9.1. Оцените сверху величину $|y(1) - y_n|$, где $y(1)$ — точное решение задачи

$$y' = \sin(xy), \quad y(0) = 0$$

в точке $x = 1$, а y_n — приближение к нему, полученное методом Эйлера с шагом $h = \frac{1}{n}$.

9.2. Продолжите серию разностных схем (9.11)–(9.15), используя формулы численного дифференцирования третьего порядка точности.

9.3. Исследуйте на устойчивость по Дальквисту:

а) явный и неявный двухшаговые разностные методы второго порядка (9.14) и (9.15);

б) явный двухшаговый метод третьего порядка (8.46);

в) явный и неявный методы, выведенные при выполнении упр.8.3.

9.4. Исследуйте на модельном уравнении $y' = py$ поведение ошибки метода Адамса–Башфорта второго порядка (8.12). Что можно сказать о его устойчивости?

9.5. Проанализируйте выражения корней характеристического уравнения для однопараметрического семейства двухшаговых методов третьего порядка (8.45), выделите из этого семейства наиболее устойчивый по Дальквисту метод. Встречался ли этот метод ранее? Как он называется?

9.6. Убедитесь, что $A(\alpha)$ -устойчивый чисто неявный метод третьего порядка (9.62) является строго устойчивым по Дальквисту.

9.7. При каком ограничении на шаг можно рассчитывать на устойчивое решение задачи

$$\begin{cases} \dot{x} = -11x + y, & x(0) = 0, \\ \dot{y} = -7x - 3y, & y(0) = 0 \end{cases}$$

методом Эйлера?

9.8. Не находя области устойчивости чисто неявного метода второго порядка (9.15), покажите, что для всех точек комплексной плоскости чисел $\mu = p/h$ выполняется условие $|\lambda_{1,2}| \leq 1$, где $\lambda_{1,2}$ — корни характеристического уравнения (9.56).

ГЛАВА 10 || МЕТОДЫ ПРИБЛИЖЕННОГО РЕШЕНИЯ КРАЕВЫХ ЗАДАЧ ДЛЯ ОБЫКНОВЕННЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ

Рассматривается несколько способов приближенного решения двухточечных линейных краевых задач для дифференциальных уравнений второго порядка.

Сначала изучаются методы, позволяющие применить накопленный арсенал методов численного решения начальных задач. Далее на основе простейших аппроксимаций производных в узлах сетки строится удобный для автоматизированных вычислений метод конечных разностей получения каркасов решений и изучается его численная устойчивость. Из приближенно-аналитических методов здесь представлены метод коллокации и метод Галёркина. Показывается, как можно подбирать требуемые этими методами системы базисных функций в случае произвольных линейных краевых условий. В развитие метода Галёркина выводятся расчетные формулы метода конечных элементов, где в качестве базисных функций используются В-сплайны. Подобно конечноразностному методу его применение сводится к решению линейных алгебраических систем с трехдиагональными матрицами коэффициентов.

§ 10.1. ПОСТАНОВКА ЗАДАЧИ. КЛАССИФИКАЦИЯ ПРИБЛИЖЕННЫХ МЕТОДОВ

Будем рассматривать двухточечные краевые задачи для обыкновенных дифференциальных уравнений второго порядка. Общий вид таких задач

$$\begin{aligned} F(x, y, y', y'') &= 0, \quad x \in [a, b], \\ \varphi_1(y(a), y'(a)) &= 0, \quad \varphi_2(y(b), y'(b)) = 0, \end{aligned} \quad (10.1)$$

где F, φ_1, φ_2 — заданные функции определенной гладкости. Наиболее употребительны и лучше всего изучены линейные краевые задачи, т.е. задачи вида (10.1), в которых F, φ_1 и φ_2 — линейные функции. Для определенности будем считать основным объектом дальнейшего изучения **линейную краевую задачу**

$$L[y] := y'' + p(x)y' + q(x)y = f(x), \quad x \in [a, b], \quad (10.2)$$

$$l_a[y] := \alpha_0 y(a) + \alpha_1 y'(a) = A, \quad (10.3)$$

$$l_b[y] := \beta_0 y(b) + \beta_1 y'(b) = B, \quad (10.4)$$

где к коэффициентам **краевых условий** (10.3), (10.4) предъявляется требование

$$|\alpha_0| + |\alpha_1| \neq 0, \quad |\beta_0| + |\beta_1| \neq 0, \quad (10.5)$$

а функции $p = p(x)$, $q = q(x)$ и $f = f(x)$ в уравнении (10.2) должны быть такими, чтобы данная задача имела единственное решение $y = y(x)$ в заданном функциональном пространстве. Краевые условия (10.3), (10.4) определяют так называемую **третью** или, иначе, **смешанную краевую задачу** для уравнения (10.2), содержащую в себе **первую** (когда $\alpha_1 = \beta_1 = 0$) или **вторую** (при $\alpha_0 = \beta_0 = 0$) краевые задачи. Оставив в стороне важный случай **периодической краевой задачи**, когда вместо (10.3), (10.4) выставляются условия $y(a) = y(b)$, $y'(a) = y'(b)$, будем конструировать методы приближенного решения смешанной задачи (10.2)–(10.4), лишь изредка выделяя случаи первой и/или второй задач, если это окажется существенным.

Как известно, точное (аналитическое) решение краевых задач вызывает еще большие трудности, чем решение задач Коши. Отсюда — повышенный интерес и большое разнообразие приближенных методов решения таких задач. По типу представления результатов приближенного решения методы можно разделить на две группы: приближенно-аналитические, дающие приближенное решение краевой задачи на отрезке $[a, b]$ в виде некоторой конкретной функции, и собственно численные или сеточные методы, дающие каркас приближенного решения на заданной на $[a, b]$ сетке^{*)}. По идейной основе приближенных методов их можно классифицировать следующим образом:

- 1) методы сведения к задаче Коши (метод пристрелки, метод дифференциальной прогонки, метод редукции);
- 2) метод конечных разностей;
- 3) метод балансов или интегро-интерполяционный метод;
- 4) метод коллокации;
- 5) проекционные методы (моментов, Галёркина);
- 6) вариационные методы (наименьших квадратов, Ритца);
- 7) проекционно-разностные методы (метод конечных элементов);
- 8) методы сведения к интегральным уравнениям Фредгольма и др.

^{*)} При этом некоторые методы первой группы также предполагают использование сетки.

Методы 4...6 из вышеперечисленных приводят к приближенному решению в виде функции заданного семейства (линейной комбинации некоторой системы линейно независимых функций), методы 1...3 и 7 генерируют таблицы численных значений приближенного решения, в методах 8 возможны варианты. Как правило, чисто сеточные методы являются более простыми и позволяют технически легко строить каркас решения на заданной сетке с наперед заданной точностью, контролируемой, например, по принципу Рунге. Однако и приближенно-аналитические методы имеют свои достоинства, одно из которых очевидно — это лаконичность функционального представления решения; другое же их достоинство состоит в том, что некоторые методы этой группы позволяют получать хорошие приближения к обобщенным решениям краевой задачи, когда она не имеет единственного решения в классическом смысле.

Ниже будут рассмотрены идеи и некоторые реализации методов 1, 2, 4, 5, 7 приведенного списка; при желании познакомиться с другими методами или с иным освещением перечисленных читатель может обратиться к другим учебным пособиям, например, к [9, 31, 55, 81].

§ 10.2. МЕТОДЫ СВЕДЕНИЯ КРАЕВЫХ ЗАДАЧ К НАЧАЛЬНЫМ

Имея в виду, что нами изучено много способов приближенного решения задач Коши для ОДУ, можно считать краевую задачу **п р и н ц и п и а л ь н о** решенной, если ее удастся преобразовать к эквивалентной начальной задаче^{*)}. Известно несколько приемов, с помощью которых можно это сделать. Эти приемы приводят к необходимости решать не одну, а несколько задач Коши разной сложности при различных ограничениях на параметры исходной задачи. Рассмотрим вкратце три таких приема.

^{*)} Автор считает уместным напомнить здесь весьма популярную в математических кругах притчу о различии подходов к решению прикладных задач двух групп людей, представителей которых будем условно называть «практик» и «теоретик».

Практику и теоретика дали по доске с забитым **п о ш л я п к у** гвоздем и инструменты: молоток и клещи, предложив вытащить гвоздь. Повозившись, оба справились с задачей (не будем уточнять, кто быстрее). Второе задание состояло в том, чтобы вытащить гвоздь, вбитый в доску **л и ш ь н а п о л о в и н у**. Практик взял клещи и вытащил гвоздь. Теоретик взял молоток и со словами: «Задача сведена к предыдущей», забил гвоздь **п о ш л я п к у**.

Метод пристрелки. Пусть требуется найти приближенное решение уравнения (в общем случае нелинейного)

$$y'' = f(x, y, y'), \quad x \in [a, b] \quad (10.6)$$

с краевыми условиями первого рода

$$y(a) = A, \quad y(b) = B. \quad (10.7)$$

Зададимся некоторым числом C_1 и будем рассматривать задачу Коши для уравнения (10.6) с начальными условиями

$$y(a) = A, \quad y'(a) = C_1.$$

Предположим, что на отрезке $[a, b]$ построено приближенное решение $y = y_1(x, C_1)$ этой начальной задачи. Сравнение значения $y_1(b, C_1)$ полученного решения в точке $x = b$ с заданным, согласно (10.7), значением B дает информацию для корректирования угла наклона касательной к решению $y = y_2(x, C_2)$ новой начальной задачи с условиями

$$y(a) = A, \quad y'(a) = C_2$$

так, чтобы уменьшить разницу $B - y_2(b, C_2)$. На этой основе могут выстраиваться те или иные стратегии варьирования задаваемых значений $y'(a) = C_1, C_2, C_3, \dots$, обеспечивающих, по возможности, наиболее быструю практическую сходимость к числу B последовательности приближений $y_k(b, C_k)$ ($k = 1, 2, 3, \dots$)^{*}. Если при заданном $\varepsilon > 0$ и некотором $k = n$ будет выполнено неравенство $|B - y_n(b, C_n)| \leq \varepsilon$, за искомое приближенное решение краевой задачи (10.6)–(10.7) принимается функция $y = y_n(x, C_n)$ (дискретная или непрерывная).

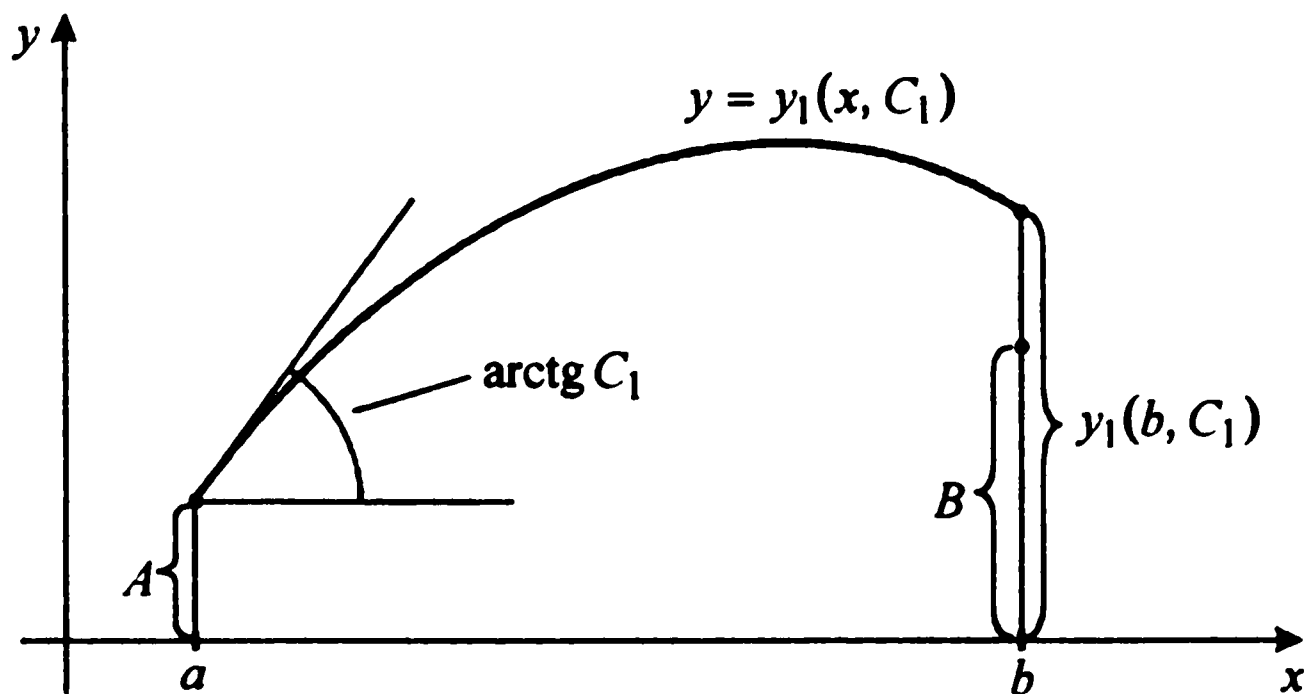


Рис. 10.1. Геометрическая интерпретация одного шага метода пристрелки

^{*} См., например, [2, 87].

Схематично описанный способ получения приближенного решения первой краевой задачи путем последовательного решения нескольких задач Коши для того же уравнения называют **методом пристрелки** (или **стрельбы**). Такое название становится естественным, если рассмотреть геометрическую (или, если угодно, баллистическую) интерпретацию метода (рис. 10.1).

Метод редукции. Будем искать решение $y = y(x)$ линейной краевой задачи (10.2)–(10.4) в виде

$$y = Cu(x) + v(x), \quad (10.8)$$

где C и $u = u(x)$, $v = v(x)$ — некоторые постоянная и функции, условия на которые будут накладываться ниже.

Сначала потребуем, чтобы функция (10.8) была решением уравнения (10.2) при любом значении постоянной C , т.е. чтобы имело место

$$L[Cu + v] = f \quad \forall C \in \mathbb{R}.$$

В силу линейности заданного в (10.2) дифференциального оператора L , последнее можно переписать в виде

$$CL[u] + L[v] = f \quad \forall C \in \mathbb{R}, \quad (10.9)$$

а так как постоянная C должна быть произвольной, то уравнение (10.9) равносильно системе

$$\begin{cases} L[u] = 0, \\ L[v] = f. \end{cases} \quad (10.10)$$

Теперь воспользуемся линейностью краевого условия (10.3). Имеем:

$$l_a[y] = l_a[Cu + v] = Cl_a[u] + l_a[v] = A \quad \forall C \in \mathbb{R}$$

$$\Leftrightarrow \begin{cases} l_a[u] = 0, \\ l_a[v] = A \end{cases} \Leftrightarrow \begin{cases} \alpha_0 u(a) + \alpha_1 u'(a) = 0, \\ \alpha_0 v(a) + \alpha_1 v'(a) = A. \end{cases} \quad (10.11)$$

$$(10.12)$$

Равенство (10.11) будет выполнено, если положить, например,

$$u(a) = \alpha_1, \quad u'(a) = -\alpha_0. \quad (10.13)$$

Для удовлетворения равенства (10.12) можно взять

$$v(a) = \frac{A}{\alpha_0}, \quad v'(a) = 0, \quad (10.14)$$

если $\alpha_0 \neq 0$, или

$$v(a) = 0, \quad v'(a) = \frac{A}{\alpha_1}, \quad (10.15)$$

если $\alpha_1 \neq 0$. Заметим, что одновременно α_0 и α_1 в нуль не обращаются, в силу условия (10.5); неоднозначность задания начальных условий для u и v связана с лишней степенью свободы: одна неизвестная функция u заменяется линейной комбинацией двух функций (u и v).

Итак, согласно (10.10), функции u и v в представлении решения (10.8) можно найти, решая уравнения

$$u'' + p(x)u' + q(x)u = 0 \quad (10.16)$$

и

$$v'' + p(x)v' + q(x)v = f(x) \quad (10.17)$$

с начальными условиями (10.13) для (10.16) и с начальными условиями (10.14) или (10.15) для (10.17), применяя какие-либо методы решения задач Коши для уравнений второго порядка (см. гл.8). Приближенное решение этих задач строится на отрезке $[a, b]$, в результате чего становятся известными, в частности, значения $u(b)$, $u'(b)$, $v(b)$, $v'(b)$. Это позволяет подобрать постоянную C так, чтобы с этим значением C и найденными функциями $u(x)$ и $v(x)$ функция $y = Cu + v$ удовлетворяла не только уравнению (10.2) и условию (10.3), но и второму условию (10.4).
Имеем:

$$l_b[y] = l_b[Cu + v] = Cl_b[u] + l_b[v] = B,$$

если

$$C = \frac{B - l_b[v]}{l_b[u]} = \frac{B - \beta_0 v(b) - \beta_1 v'(b)}{\beta_0 u(b) + \beta_1 u'(b)}. \quad (10.18)$$

Замечание 10.1. Если знаменатель выражения C в (10.18) окажется равным нулю, это будет означать, что однородная краевая задача

$$L[u] = 0, \quad l_a[u] = 0, \quad l_b[u] = 0$$

имеет нетривиальное решение $u(x)$, что, в свою очередь, служит признаком вырожденности исходной задачи (10.2)–(10.4).

Метод дифференциальной прогонки. При выводе метода прогонки решения линейных алгебраических систем с трехдиагональными матрицами коэффициентов, т.е. для трехточечных разностных уравнений второго порядка, связывающих три

соседние компоненты вектора неизвестных системы, в [18] мы исходили из предположения о наличии связей между каждыми двумя соседними неизвестными и получали рекуррентные формулы для коэффициентов таких связей. Проводя параллель между линейными дифференциальными и линейными разностными уравнениями, предположим, что существуют такие две функции $\delta = \delta(x)$ и $\gamma = \gamma(x)$, с помощью которых решение $y = y(x)$ данного линейного дифференциального уравнения второго порядка (10.2) может быть представлено как решение уравнения первого порядка

$$y' = \delta(x)y + \gamma(x). \quad (10.19)$$

Продифференцируем это равенство и подставим выражение второй производной $y'' = \delta'y + \delta y' + \gamma'$ в исходное уравнение (10.2). Имеем

$$\delta'y + \delta y' + \gamma' + py' + qy = f,$$

откуда, выразив y' , получаем равенство типа (10.19):

$$y' = -\frac{\delta' + q}{\delta + p}y + \frac{f - \gamma'}{\delta + p}. \quad (10.20)$$

Уравнения (10.20) и (10.19) можно отождествлять при условии, что

$$\begin{cases} \delta = -\frac{\delta' + q}{\delta + p}, \\ \gamma = \frac{f - \gamma'}{\delta + p}. \end{cases}$$

Следовательно, функции $\delta = \delta(x)$ и $\gamma = \gamma(x)$ должны удовлетворять системе дифференциальных уравнений первого порядка

$$\begin{cases} \delta' = -p(x)\delta - \delta^2 - q(x), \end{cases} \quad (10.21)$$

$$\begin{cases} \gamma' = -(p(x) + \delta)\gamma + f(x). \end{cases} \quad (10.22)$$

Чтобы найти начальные условия для уравнений системы (10.21), (10.22), учтем, что решение $y(x)$ в точке a должно удовлетворять уравнению (10.19), т.е. должно выполняться равенство

$$y'(a) = \delta(a)y(a) + \gamma(a). \quad (10.23)$$

С другой стороны, в предположении, что $\alpha_1 \neq 0$, из краевого условия (10.3) для того же решения в точке a имеем

$$y'(a) = -\frac{\alpha_0}{\alpha_1} y(a) + \frac{A}{\alpha_1}. \quad (10.24)$$

Сравнение (10.23) с (10.24) приводит к начальным условиям

$$\begin{cases} \delta(a) = -\frac{\alpha_0}{\alpha_1}, \end{cases} \quad (10.25)$$

$$\begin{cases} \gamma(a) = \frac{A}{\alpha_1}. \end{cases} \quad (10.26)$$

Решив на отрезке $[a, b]$ задачу Коши для системы (10.21), (10.22) с начальными условиями (10.25), (10.26), получим требуемые для уравнения (10.19) функции $\delta(x)$ и $\gamma(x)$. Для нахождения частного решения $y(x)$ уравнения (10.19), служащего одновременно решением данной краевой задачи, аналогично предыдущему запишем равенство

$$y'(b) = \delta(b)y(b) + \gamma(b)$$

и подставим это выражение $y'(b)$ в краевое условие (10.4):

$$\beta_0 y(b) + \beta_1 \delta(b)y(b) + \beta_1 \gamma(b) = B.$$

Отсюда получаем привязку решения $y(x)$ в точке b :

$$y(b) = \frac{B - \beta_1 \gamma(b)}{\beta_0 + \beta_1 \delta(b)} \quad (10.27)$$

(если $\beta_0 + \beta_1 \delta(b) \neq 0$). Остается решить задачу Коши (10.19), (10.27) от точки b к точке a .

Таким образом, **метод дифференциальной прогонки** решения линейной краевой задачи (10.2)–(10.4) заключается в решении трех начальных задач для дифференциальных уравнений первого порядка: сначала параллельно (или последовательно в записанном порядке) решаются уравнения (10.21), (10.22) с начальными условиями (10.25), (10.26) в прямом направлении от a к b , затем в обратном направлении от b к a решается уравнение (10.19) с начальным условием (10.27).

При численной реализации этого метода следует обратить внимание на необходимость согласования расчетных сеток решения начальных задач в одном и в другом направлениях так, чтобы имелась возможность получения решения последнего уравнения (10.19) с нужной точностью при условии, что в процессе численного интегрирования вспомогательных уравнений

(10.21), (10.22) находятся лишь каркасы функций $\delta(x)$ и $\gamma(x)$ на своих сетках.

Замечание 10.2. Если не выполняется сделанное выше предположение $\alpha_1 \neq 0$, то вместо описанной дифференциальной прогонки, которую естественно назвать *левой*, можно применить *правую* прогонку, предполагая $\beta_1 \neq 0$. Для этого достаточно начальные условия для нахождения функций $\delta(x)$ и $\gamma(x)$ вывести из краевого условия (10.4), а для $\gamma(x)$ — из (10.3). В результате придется строить вспомогательные решения, т.е. решать систему (10.21), (10.22) в направлении от b к a , а окончательное решение $\gamma(x)$ находить из уравнения (10.19) в прямом направлении от a к b . Очевидно, что когда $\alpha_1 = \beta_1 = 0$, т.е. в случае первой краевой задачи, рассмотренный выше метод дифференциальной прогонки неприменим (можно сказать, что они дополняют друг друга с методом пристрелки).

Имеются и другие способы сведения краевых задач к начальным (см., например, [48]).

§ 10.3. МЕТОД КОНЕЧНЫХ РАЗНОСТЕЙ

Идея *метода конечных разностей* (МКР^{*}) решения краевых задач весьма проста и видна уже из самого названия: вместо производных в дифференциальном уравнении используются их конечноразностные аппроксимации. Эта идея уже применялась нами в § 9.2 при построении простейших разностных схем решения задачи Коши. В отличие от того случая, где дискретным аналогом начальной дифференциальной задачи служила начальная задача для разностного уравнения, что позволяло последовательно, шаг за шагом, находить компоненты каркаса приближенного решения, здесь, т.е. для краевой дифференциальной задачи, будет получаться краевая же задача для разностного уравнения, а это ведет к необходимости решать алгебраическую систему. Ясно, что при построении дискретных аппроксимаций краевых дифференциальных задач нужно стремиться увязать две, возможно, противоречивые цели: хорошее качество аппроксимации и эффективное и устойчивое решение получающихся при этом алгебраических систем.

Рассмотрим наиболее типичные воплощения указанной идеи МКР для линейной краевой задачи (10.2)–(10.4).

^{*} Зарубежная аббревиатура FDM — от англ. finite difference method.

Сначала вводим на отрезке $[a, b]$ *сетку* с шагом $h = \frac{b-a}{n}$:

$$\omega_h := \{x_i \mid x_i = x_0 + ih; \quad i = 0, 1, \dots, n; \quad x_0 := a, x_n := b\}.$$

На этой сетке определяются *сеточные функции*

$$p_i := p(x_i), \quad q_i := q(x_i), \quad f_i := f(x_i), \quad (10.28)$$

отвечающие функциональным коэффициентам данного дифференциального уравнения (10.2). Считая $y(x)$ точным решением данной краевой задачи (10.2)–(10.4), через

$$y_i \approx y(x_i) \quad (10.29)$$

будем обозначать i -ю компоненту искомого каркаса приближенного решения $y_n(x) \approx y(x)$.

Фиксируя в уравнении (10.2) $x = x_i$, с учетом обозначений (10.28) приходим к равенствам

$$y''(x_i) + p_i y'(x_i) + q_i y(x_i) = f_i, \quad (10.30)$$

где целая переменная i может принимать значения от 0 до n по числу узлов сетки, а под $y(x_i)$, $y'(x_i)$, $y''(x_i)$ понимаются значения точного решения $y(x)$ и его производных в i -ом узле. В каждом внутреннем узле сетки ω_h , т.е. при $i = 1, 2, \dots, n-1$, значения производных аппроксимируем конечноразностными отношениями по симметричным формулам второго порядка точности (см. (6.18) и (6.20)):

$$y'(x_i) = \frac{y(x_{i+1}) - y(x_{i-1}))}{2h} + O(h^2),$$

$$y''(x_i) = \frac{y(x_{i+1}) - 2y(x_i) + y(x_{i-1}))}{h^2} + O(h^2).$$

В результате подстановки последних в равенства (10.30) при $i = 1, 2, \dots, n-1$ получаем

$$\begin{aligned} \frac{y(x_{i+1}) - 2y(x_i) + y(x_{i-1}))}{h^2} + p_i \frac{y(x_{i+1}) - y(x_{i-1}))}{2h} + \\ + q_i y(x_i) = f_i + O(h^2). \end{aligned}$$

Это уравнение служит точным отражением дифференциальной связи (10.2), но в нем имеется неопределенное слагаемое $O(h^2)$.

Отбрасывая его, приходим к разностному уравнению относительно приближенных значений решения (обозначения которых соответствуют (10.29)):

$$\frac{y_{i+1} - 2y_i + y_{i-1}}{h^2} + p_i \frac{y_{i+1} - y_{i-1}}{2h} + q_i y_i = f_i. \quad (10.31)$$

После приведения подобных членов в (10.31) получаем стандартное трехточечное разностное уравнение второго порядка

$$(1 + \frac{h}{2} p_i) y_{i+1} - (2 - h^2 q_i) y_i + (1 - \frac{h}{2} p_i) y_{i-1} = h^2 f_i, \quad (10.32)$$

где $i = 1, 2, \dots, n-1$.

Интерпретируя (10.32) как компактную запись системы линейных алгебраических уравнений с трехдиагональной матрицей коэффициентов, видим, что число уравнений в ней $n-1$, в то время как неизвестных — $n+1$: $y_0, y_1, \dots, y_n$. Два недостающие уравнения этой системы (или, в другой терминологии, краевые условия разностного уравнения (10.32)) следует получить на основе краевых условий (10.3), (10.4) данной задачи.

Будем рассматривать два варианта аппроксимации входящих в краевые условия значений первой производной решения в точках $a = x_0$ и $b = x_n$. Именно, здесь идет речь о несимметричных аппроксимациях первого и второго порядков точности (достоинства и недостатки каждого из этих вариантов будут обсуждены позже).

Согласно формулам (6.15), (6.26), (6.14) и (6.27), можно записать равенства

$$y'(a) = \frac{y(x_1) - y(x_0)}{h} + O(h) = \frac{-3y(x_0) + 4y(x_1) - y(x_2)}{2h} + O(h^2),$$

$$y'(b) = \frac{y(x_n) - y(x_{n-1})}{h} + O(h) = \frac{y(x_{n-2}) - 4y(x_{n-1}) + 3y(x_n)}{2h} + O(h^2).$$

В первом варианте, при аппроксимации $y'(a)$ и $y'(b)$ в (10.3) и (10.4) двухточечными разностными отношениями первого порядка, имеем

$$\alpha_0 y(x_0) + \alpha_1 \frac{y(x_1) - y(x_0)}{h} + O(h) = A,$$

$$\beta_0 y(x_n) + \beta_1 \frac{y(x_n) - y(x_{n-1})}{h} + O(h) = B.$$

Отсюда после отбрасывания слагаемого $O(h)$ (с заменой $y(x_i)$ на y_i) и упрощения получаем краевые условия для разностного уравнения (10.32), иначе, нулевое и n -е уравнения

$$(h\alpha_0 - \alpha_1)y_0 + \alpha_1 y_1 = Ah \quad (10.33)$$

и

$$-\beta_1 y_{n-1} + (h\beta_0 + \beta_1)y_n = Bh \quad (10.34)$$

системы алгебраических уравнений, задаваемой равенством (10.32).

Во втором варианте имеем аналогично

$$\alpha_0 y(x_0) + \alpha_1 \frac{-3y(x_0) + 4y(x_1) - y(x_2)}{2h} + O(h^2) = A,$$

$$\beta_0 y(x_n) + \beta_1 \frac{y(x_{n-2}) - 4y(x_{n-1}) + 3y(x_n)}{2h} + O(h^2) = B,$$

откуда следуют дополнительные к (10.32) связи между тремя первыми неизвестными

$$(2h\alpha_0 - 3\alpha_1)y_0 + 4\alpha_1 y_1 - \alpha_1 y_2 = 2Ah \quad (10.35)$$

и тремя последними

$$\beta_1 y_{n-2} - 4\beta_1 y_{n-1} + (2h\beta_0 + 3\beta_1)y_n = 2Bh. \quad (10.36)$$

Сравним теперь два рассматриваемых варианта. В первом из них СЛАУ, образованная уравнениями (10.33), (10.32), (10.34), имеет трехдиагональную матрицу коэффициентов, и к ней сразу можно применить высокоэффективный метод прогонки, который уже упоминался в §§ 4.2, 4.3 в связи с кубической и с квадратичной сплайн-интерполяцией. Во втором случае соответствующая методу прогонки структура матрицы коэффициентов СЛАУ еще должна быть создана, для чего нужно из уравнения (10.35) с помощью уравнения, получающегося из (10.32) при $i = 1$, исключить неизвестное y_2 , а из (10.36) с помощью (10.32) при $i = n - 1$ исключить y_{n-2} . Усложнения, сопутствующие второму варианту, могут быть оправданы тем, что в этом случае исходная дифференциальная краевая задача полностью аппроксимируется алгебраической системой относительно компонент каркаса решения с точностью $O(h^2)$, в то время как о первом варианте такого сказать нельзя (если, конечно, речь не идет о первой краевой задаче, т.е. о случае $\alpha_1 = \beta_1 = 0$, когда в аппроксимации краевых условий вообще нет нужды). Однако в поисках компромисса между качеством аппроксимации и численной

устойчивостью при решении конкретных задач первый вариант может оказаться и более предпочтительным.

Остановимся теперь на вопросе устойчивости построенной конечноразностной схемы решения краевой задачи (10.2)–(10.4). Как уже отмечалось ранее, эту устойчивость можно связать с устойчивостью метода прогонки, что, в свою очередь, можно гарантировать, когда матрица коэффициентов имеет свойство диагонального преобладания [18]. Посмотрим с этой точки зрения на i -е «внутреннее» уравнение системы, т.е. на уравнение (10.32).

Условие диагонального преобладания для (10.32) означает, что должно выполняться неравенство

$$\left|2 - h^2 q_i\right| > \left|1 + \frac{h}{2} p_i\right| + \left|1 - \frac{h}{2} p_i\right| \quad \forall i \in \{1, 2, \dots, n-1\}. \quad (10.37)$$

Рассмотрим, что представляет собой правая часть этого неравенства. Раскрывая модули, имеем

$$\left|1 + \frac{h}{2} p_i\right| + \left|1 - \frac{h}{2} p_i\right| = \begin{cases} -hp_i, & \text{если } hp_i < 2, \\ 2, & \text{если } |hp_i| \leq 2, \\ hp_i, & \text{если } hp_i > 2. \end{cases}$$

Следовательно, правую часть неравенства (10.37) как функцию переменной hp_i (считая ее изменяющейся непрерывно) в условных координатах можно представить в виде графика, изображенного на рис.10.2.

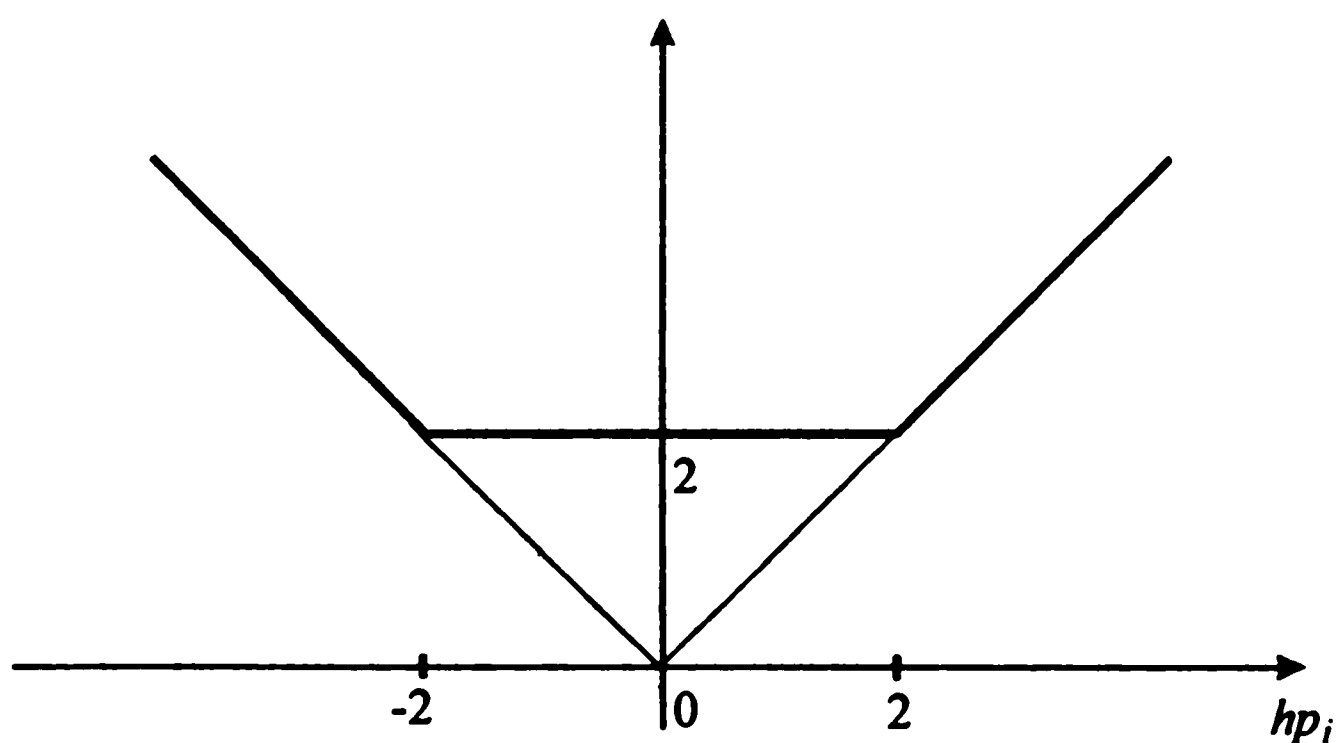


Рис. 10.2. Условный график правой части неравенства (10.37)

Так как левая часть неравенства (10.37) при $q_i > 0$ и малых $h > 0$ (малость h нужна из требований аппроксимации) меньше 2, то на устойчивость прогонки можно рассчитывать лишь в случае, когда $q(x) < 0$. При этом имеет место

$$|2 - h^2 q_i| = 2 - h^2 q_i > 2 \quad \forall h.$$

Чтобы в таком случае неравенство (10.37) выполнялось при любых $p(x)$, для правой его части считаем допустимым только значение 2 (т.е. используем горизонтальную часть графика на рис.10.2). Отсюда получаем ограничение

$$|hp_i| \leq 2,$$

означающее, что устойчивость прогонки можно гарантировать при условии, что шаг дискретизации h удовлетворяет неравенству

$$h \leq \frac{2}{|p_i|} \quad \forall i \in \{1, 2, \dots, n-1\}.$$

Усиливая это неравенство и используя сформулированное в конце § 9.1 утверждение «Аппроксимация плюс устойчивость дает сходимость», приходим к заключению, что *если в дифференциальном уравнении (10.2)*

$$q(x) < 0 \quad \forall x \in [a, b], \quad (10.38)$$

а в определяющем МКР разностном уравнении (10.32)

$$h \leq \frac{2}{\max_{x \in [a, b]} |p(x)|},$$

то МКР сходится (по крайней мере, к решению первой краевой задачи, т.е. когда в (10.3), (10.4) $\alpha_1 = \beta_1 = 0$; в других случаях требуется более детальный анализ).

Наличие ограничения на шаг h в методе конечных разностей второго порядка (10.32) характеризует его как условно устойчивый метод. Если отказаться от аппроксимации всех производных с порядком $O(h^2)$ и использовать в роли $y'(x_i)$ правые или левые разностные отношения первого порядка точности, связывая их выбор со знаком p_i , а именно, рассматривая вместо

(10.31) разностное уравнение

$$\frac{y_{i+1} - 2y_i + y_{i-1}}{h^2} + p_i \left[\begin{array}{l} \frac{y_{i+1} - y_i}{h}, \text{ если } p_i > 0 \\ \frac{y_i - y_{i-1}}{h}, \text{ если } p_i < 0 \end{array} \right] + q_i y_i = f_i$$

при $i = 1, 2, \dots, n-1$, приходим к конечноразностному методу

$$\begin{cases} y_{i-1} - (2 + hp_i - h^2 q_i) y_i + (1 + hp_i) y_{i+1} = h^2 f_i, & \text{если } p_i > 0, \\ (1 - hp_i) y_{i-1} - (2 - hp_i - h^2 q_i) y_i + y_{i+1} = h^2 f_i, & \text{если } p_i < 0, \end{cases} \quad (10.39)$$

имеющему первый порядок точности независимо от точности аппроксимации краевых условий.

Легко видеть, что при условии (10.38) диагональное преобладание в методе (10.39) будет при любой величине шага $h > 0$. Отсюда следует его *безусловная устойчивость*, правда в ущерб точности; последнее означает необходимость проведения вычислений с более мелким шагом для доведения погрешности решения до некоторой фиксированной малой величины, чем это требует метод второго порядка (10.32), если они оба одновременно применимы. Конечноразностный метод (10.39) широко используется при решении задач динамики жидкости и газов и называется *upwind*- или *противопотоковым методом*.

§ 10.4. МЕТОД КОЛЛОКАЦИИ

Будем искать приближенное решение линейной краевой задачи (10.2)–(10.4) в виде функции

$$y_n(x) := \varphi_0(x) + \sum_{i=1}^n c_i \varphi_i(x), \quad (10.40)$$

где, определяемые на отрезке $[a, b]$ **базисные функции** $\varphi_i(x)$ ($i = 1, 2, \dots, n$) и дополнительная функция $\varphi_0(x)$ должны быть дважды дифференцируемыми и попарно линейно независимыми. Кроме того, функция $\varphi_0(x)$ должна удовлетворять данным краевым условиям (10.3), (10.4), а функции $\varphi_i(x)$ при $i = 1, 2, \dots, n$ — соответствующим однородным краевым условиям, т.е. должны выполняться равенства

$$\left. \begin{array}{l} \alpha_0 \varphi_i(a) + \alpha_1 \varphi_i'(a) = 0, \\ \beta_0 \varphi_i(b) + \beta_1 \varphi_i'(b) = 0 \end{array} \right\} \quad \forall i \in \{i = 1, 2, \dots, n\}. \quad (10.41)$$

В таком случае функция $y_n(x)$, определяемая выражением (10.40), при любых значениях коэффициентов c_i гарантированно удовлетворяет краевым условиям (10.3), (10.4). Действительно, например, в точке $x = a$ имеем

$$\begin{aligned}\alpha_0 y_n(a) + \alpha_1 y'_n(a) &= \\ &= \alpha_0 \varphi_0(a) + \alpha_1 \varphi'_0(a) + \alpha_0 \sum_{i=1}^n c_i \varphi_i(a) + \alpha_1 \sum_{i=1}^n c_i \varphi'_i(a) = \\ &= A + \sum_{i=1}^n c_i [\alpha_0 \varphi_i(a) + \alpha_1 \varphi'_i(a)] = A;\end{aligned}$$

аналогично при $x = b$ с помощью (10.4), (10.40) и (10.41) проверяется справедливость равенства

$$\beta_0 y_n(b) + \beta_1 y'_n(b) = B.$$

Представление приближенного решения $y_n(x)$, подобное (10.40), характерно для многих приближенно-аналитических методов решения краевых задач (возможны вариации требований к базисным функциям); главное их различие состоит в том, на какой основе ищутся коэффициенты c_i в линейной комбинации базисных функций $\varphi_i(x)$ выражения (10.40).

В *методе коллокации*^{*} коэффициенты c_i в представлении (10.40) приближенного решения $y_n(x)$ подбираются так, чтобы в узлах коллокации x_i таких, что

$$a < x_1 < x_2 < \dots < x_n < b$$

(не обязательно равноотстоящих, но строго внутренних точках отрезка $[a, b]$), значения $y_n(x_i)$ приближенного решения были согласованы с точными значениями $y(x_i)$.

Поскольку точное решение $y(x)$ задачи (10.2)–(10.4) неизвестно, согласование $y_n(x)$ и $y(x)$ в узлах коллокации x_i проводим подстановкой $y_n(x)$ в уравнение (10.2). Имеем равенство

$$y_n''(x_i) + p(x_i)y_n'(x_i) + q(x_i)y_n(x_i) = f(x_i), \quad (10.42)$$

которое, в силу выставляемого требования согласования $y_n(x_i)$ с $y(x_i)$, считаем точным при каждом $i \in \{1, 2, \dots, n\}$. Продифференцировав дважды функцию $y_n(x)$ в представлении (10.40), от

^{*} Collocatio (лат.) — размещение, расстановка. В книге [41] такой метод называют *интерполяционным методом* или *методом совпадений*.

равенства (10.42) переходим к равенству

$$\begin{aligned} \sum_{j=1}^n c_j \varphi_j''(x_i) + p_i \sum_{j=1}^n c_j \varphi_j'(x_i) + q_i \sum_{j=1}^n c_j \varphi_j(x_i) = \\ = f_i - \varphi_0''(x_i) - p_i \varphi_0'(x_i) - q_i \varphi_0(x_i), \end{aligned} \quad (10.43)$$

где p_i, q_i, f_i соответствуют обозначениям (10.28) сеточных функций. Положим

$$a_{ij} := \varphi_j''(x_i) + p_i \varphi_j'(x_i) + q_i \varphi_j(x_i), \quad (10.44)$$

$$b_i := f_i - \varphi_0''(x_i) - p_i \varphi_0'(x_i) - q_i \varphi_0(x_i). \quad (10.45)$$

Тогда (10.43) приобретает стандартный вид линейной алгебраической системы

$$\sum_{j=1}^n a_{ij} c_j = b_i, \quad i = 1, 2, \dots, n \quad (10.46)$$

относительно коэффициентов $c_1, c_2, \dots, c_n$. Решив эту систему каким-нибудь стандартным методом и подставив найденные коэффициенты c_i в выражение (10.40), получаем приближенное решение $y_n(x)$.

Успех применения метода коллокации к задаче (10.2)–(10.4), впрочем, как и других приближенно-аналитических методов, сильно зависит от удачного выбора базисных функций $\varphi_i(x)$ в представлении приближенного решения (10.40). В конкретных задачах выбор таких функций, по возможности, должен опираться на априорные или эмпирические сведения о решении. В отсутствие таковых, т.е. в рассматриваемом абстрактном случае, для смешанной краевой задачи (10.2)–(10.4) можно предложить, например, следующий набор базисных функций.

В качестве φ_0 возьмем линейную функцию

$$\varphi_0(x) = \delta + \gamma x, \quad (10.47)$$

коэффициенты которой подберем так, чтобы она удовлетворяла неоднородным краевым условиям (10.3), (10.4), т.е. из линейной алгебраической системы

$$\begin{cases} \alpha_0 \delta + (\alpha_0 a + \alpha_1) \gamma = A, \\ \beta_0 \delta + (\beta_0 b + \beta_1) \gamma = B. \end{cases} \quad (10.48)$$

Функции $\varphi_i(x)$ при $i = 1, 2, \dots, n$ можно взять однопараметрическими вида

$$\varphi_i(x) = \gamma_i(x-a)^i + (x-a)^{i+1}, \quad (10.49)$$

если в (10.3) $\alpha_1 = 0$, или вида

$$\varphi_i(x) = \gamma_i(x-a)^{i+1} + (x-a)^{i+2} \quad (10.50)$$

в самом общем случае. Очевидно, что при любых γ_i эти функции удовлетворяют первому из требуемых равенств (10.41)^{*}, а если зафиксировать

$$\gamma_i = -\frac{\beta_0(b-a)^2 + (i+1)\beta_1(b-a)}{\beta_0(b-a) + i\beta_1} \quad (10.51)$$

в выражении (10.49) и

$$\gamma_i = -\frac{\beta_0(b-a)^2 + (i+2)\beta_1(b-a)}{\beta_0(b-a) + (i+1)\beta_1} \quad (10.52)$$

в (10.50), то они будут подчиняться и второму из этих равенств. Следовательно, можно рассчитывать, что с такими базисными функциями при найденных методом коллокации (или каким-либо другим методом) коэффициентах c_i определенная посредством (10.40) функция $y_n(x)$ будет удовлетворять краевым условиям и может служить приближенным решением данной краевой задачи (10.2)–(10.4).

Проблема формального выбора базисных функций φ_i значительно упрощается в случае, когда в задаче (10.2)–(10.4) фигурируют однородные краевые условия первого рода, т.е. когда

$$y(a) = 0, \quad y(b) = 0. \quad (10.53)$$

В такой ситуации в выражении (10.40) не нужна функция φ_0 , а в роли φ_i ($i = 1, 2, \dots, n$) могут выступать, например, функции

$$\varphi_i(x) = (x-a)^i(b-x)$$

или

$$\varphi_i(x) = \sin \frac{i(x-a)}{b-a} \pi.$$

^{*} Если $\alpha_1 \neq 0$, то первому из равенств (10.41) не удовлетворяет одна функция семейства (10.49), а именно $\varphi_1(x)$.

К этому случаю, т.е. к условиям (10.53), легко свести более общий случай неоднородных краевых условий первого рода

$$y(a) = A, \quad y(b) = B. \quad (10.54)$$

С этой целью достаточно сделать линейную замену (линейный сдвиг)

$$y = u + v, \quad \text{где} \quad v := A + \frac{B - A}{b - a}(x - a).$$

Дважды дифференцируя эту функцию y и подставляя результаты в уравнение (10.2), от задачи (10.2), (10.54) приходим к краевой задаче с однородными краевыми условиями относительно новой переменной u :

$$u'' + p(x)u' + q(x)u = f(x) - \frac{B - A}{b - a} p(x) - vq(x), \quad x \in [a, b],$$

$$u(a) = 0, \quad u(b) = 0.$$

Пример 10.1. Применим метод коллокации для приближенного представления решения краевой задачи

$$x^4 y'' + x^6 y' - x^5 y = 6 - 3x^3, \quad x \in [1, 2], \quad (10.55)$$

$$y(1) = 1, \quad 3y(2) + y'(2) = 0.5.$$

Составив систему (10.48) относительно коэффициентов δ и γ функции $\varphi_0(x)$ вида (10.47)

$$\begin{cases} \delta + \gamma = 1, \\ 3\delta + 7\gamma = 0.5, \end{cases}$$

находим удовлетворяющую данным краевым условиям линейную функцию $\varphi_0(x) = \frac{13}{8} - \frac{5}{8}x$. Ограничиваясь одной базисной функцией $\varphi_1(x)$ вида (10.49), по формуле (10.51) вычисляем коэффициент

$$\gamma_1 = -\frac{3(2-1)^2 + 2(2-1)}{3(2-1) + 1 \cdot 1} = -\frac{5}{4},$$

подстановка которого в (10.49) при $i=1$ дает базисную функцию

$$\varphi_1(x) = (x-1)^2 - \frac{5}{4}(x-1).$$

Возьмем в качестве узла коллокации середину рассматриваемого промежутка — точку $x_1 = \frac{3}{2}$ — и потребуем, чтобы функция

$$y_1(x) = \varphi_0(x) + c_1 \varphi_1(x)$$

удовлетворяла в этой точке заданному дифференциальному уравнению. Подставив в него

$$y_1\left(\frac{3}{2}\right) = \frac{11}{16} - \frac{3}{8}c_1, \quad y'_1 = -\frac{5}{8} - \frac{1}{4}c_1 \quad \text{и} \quad y''_1 = 2c_1,$$

получаем $c_1 = \frac{701}{864}$. Таким образом, простейшая коллокация с одним узлом приводит к приближению решения данной краевой задачи (10.55) квадратичной функцией

$$y_1(x) = 1 - \frac{5}{8}(x - 1) + \frac{701}{864}(x - 1)\left(x - \frac{9}{4}\right)$$

(ее точное решение $y(x) = \frac{1}{x^2}$).

Замечание 10.3. Приближенное решение $y_n(x)$ (10.40), согласно идее метода коллокации, в точках коллокации $x_i \in (a, b)$ должно удовлетворять данному дифференциальному уравнению (10.2). Казалось бы, в этих точках оно должно совпадать с точным решением $y(x)$ данной краевой задачи (10.2)–(10.4). Однако это не так, в чем легко убедиться, сравнив полученное в примере 10.1 приближенное решение в узле коллокации $x_1 = \frac{3}{2}$, равное $y_1\left(\frac{3}{2}\right) \approx 0.38$, с точным значением решения $y\left(\frac{3}{2}\right) = 0.4$. (См. также сравнение графиков $y_1(x)$ и $y(x)$ далее на рис.10.3.)

Замечание 10.4. Нет принципиальных препятствий к привлечению метода коллокации для приближенного решения нелинейных краевых задач. Трудности на этом пути ожидают при подборе базисных функций φ_i , удовлетворяющих краевым условиям (10.1) в случае их нелинейности, и в необходимости решать нелинейные системы при отыскании коэффициентов c_i того же представления (10.40) приближенного решения. Например, если данное дифференциальное уравнение имеет вид

$$y'' = f(x, y, y'),$$

то коэффициенты $c_1, c_2, \dots, c_n$ в представлении $y_n(x)$ вида (10.40) должны находиться из системы

$$\sum_{j=0}^n c_j \varphi_j''(x_i) = f\left(x_i, \sum_{j=0}^n c_j \varphi_j(x_i), \sum_{j=0}^n c_j \varphi_j'(x_i)\right),$$

где $i = 1, 2, \dots, n$, а $c_0 := 1$ (или $c_0 := 0$, если это дифференциальное уравнение сопровождается краевыми условиями первого рода).

§ 10.5. МЕТОД ГАЛЁРКИНА

Чтобы лучше понять основную идею *проекционных методов*, наиболее ярким представителем которых является метод Галёркина, ненадолго отвлечемся от рассматриваемой краевой задачи и сделаем небольшой экскурс в функциональный анализ.

Пусть L — некоторый линейный оператор, действующий в *гильбертовом пространстве* H , т.е. в полном нормированном пространстве со скалярным произведением $(\cdot, \cdot)$. Стоит задача приближенного решения *операторного уравнения*

$$Ly = f, \quad (10.56)$$

т.е. задача отыскания некоторого приближения к неизвестному элементу $y \in H$, соответствующему заданному элементу $f \in H$.

Пусть, далее, $\{\varphi_i\}_{i=1}^{\infty}$ — некоторая полная замкнутая система линейно независимых элементов из H . Ее n первых элементов $\varphi_1, \dots, \varphi_n$ выделяют в H конечномерное подпространство H_n , в котором и ищется приближенное решение уравнения (10.56):

$$y_n := \sum_{i=1}^n c_i \varphi_i. \quad (10.57)$$

Для удобства будем считать, что элемент Ly_n принадлежит тому же подпространству H_n . Тогда к тривиальному равенству

$$f = Ly_n + (f - Ly_n)$$

можно применить одну из центральных теорем теории гильбертовых пространств, согласно которой *любой элемент гильбертова пространства может быть представлен в виде суммы определенного элемента подпространства (проекции данного элемента на подпространство) и определенного элемента пространства, ортогонального к выбранному подпространству (реализующего расстояние от исходного элемента до его проекции)* [23, 76]^{*}.

Принадлежность элемента $f - Ly_n$ ортогональному к H_n подпространству $H_n^{\perp}$ означает, что он ортогонален каждому

^{*} Представьте себе, например, точку трехмерного пространства в декартовой системе координат, спроектируйте ее радиус-вектор на одну из координатных плоскостей; полученный при этом прямоугольный векторный треугольник может служить простейшей геометрической моделью этой теоремы.

элементу φ_i , входящему в базис пространства H_n . Таким образом, при любом $i = 1, 2, \dots, n$ имеем:

$$(f - Ly_n) \perp \varphi_i \Leftrightarrow (f - Ly_n, \varphi_i) = 0 \Leftrightarrow (Ly_n, \varphi_i) = (f, \varphi_i).$$

Подставляя сюда выражение y_n (10.57) и пользуясь простейшими свойствами скалярного произведения, получаем:

$$\begin{aligned} \left(L \sum_{j=1}^n c_j \varphi_j, \varphi_i \right) &= (f, \varphi_i) \Leftrightarrow \\ &\Leftrightarrow \sum_{j=1}^n (L\varphi_j, \varphi_i) c_j = (f, \varphi_i). \end{aligned} \quad (10.58)$$

Итак, *метод Галёркина*^{*)} приближенного решения операторного уравнения (10.56) сводится к нахождению коэффициентов $c_1, \dots, c_n$ линейной комбинации некоторых задаваемых определенным образом линейно независимых функций^{**)} $\varphi_1, \dots, \varphi_n$ так, чтобы эти коэффициенты удовлетворяли линейной системе

$$\sum_{j=1}^n a_{ij} c_j = d_i, \quad i = 1, 2, \dots, n, \quad (10.59)$$

где

$$a_{ij} := (L\varphi_j, \varphi_i), \quad d_i := (f, \varphi_i). \quad (10.60)$$

Вернемся к краевой задаче (10.2)–(10.4). Естественным для ее рассмотрения пространством является гильбертово пространство $L_2[a, b]$ функций, интегрируемых на отрезке $[a, b]$ с квадратом. Скалярное произведение здесь определяется равенством

$$(u, v) = \int_a^b u(x)v(x)dx. \quad (10.61)$$

Как и в предыдущем параграфе, будем искать приближенное решение задачи (10.2)–(10.4) в виде задаваемой формулой

^{*)} Галёркин Борис Григорьевич (1871–1945) — русский инженер и ученый в области теории упругости.

^{**)} Будем называть их *базисными функциями*, хотя, может быть, чаще в методах проекционной природы их называют *координатными функциями*.

(10.40) функции $y_n(x)$ такой, чтобы она удовлетворяла данным краевым условиям, а это будет, как мы уже знаем, если

$$l_a[\varphi_0] = A, \quad l_b[\varphi_0] = B,$$

а

$$l_a[\varphi_i] = 0, \quad l_b[\varphi_i] = 0 \quad \forall i \in \{1, 2, \dots, n\}.$$

Подставляя $y_n(x)$ в данное уравнение (10.2), в силу линейности определенного там дифференциального оператора $L[y]$, имеем:

$$L\left[\varphi_0 + \sum_{i=1}^n c_i \varphi_i\right] = f \Leftrightarrow L\left[\sum_{i=1}^n c_i \varphi_i\right] = f - L[\varphi_0].$$

Следовательно, в таком случае, наличие дополнительного слагаемого φ_0 в представлении приближенного решения (10.40) вместо представления (10.57) означает, что равенство (10.58) нужно переписать в виде

$$\left(L \sum_{j=1}^n c_j \varphi_j, \varphi_i\right) = (f - L\varphi_0, \varphi_i),$$

а это, в свою очередь, вносит поправку в выражение d_i , зафиксированное равенством (10.60). Согласно этому замечанию и определению (10.61) скалярного произведения, находим выражение свободного члена d_i i -го уравнения системы (10.59):

$$\begin{aligned} d_i &= (f - L[\varphi_0], \varphi_i) = \\ &= \int_a^b [f(x) - \varphi_0''(x) - p(x)\varphi_0'(x) - q(x)\varphi_0(x)]\varphi_i(x)dx. \end{aligned} \quad (10.62)$$

Теперь будем выражать коэффициенты a_{ij} СЛАУ (10.59) в соответствии с записанной в (10.60) формулой и скалярным произведением (10.61):

$$\begin{aligned} a_{ij} &= (L\varphi_j, \varphi_i) = \int_a^b [\varphi_j''(x) + p(x)\varphi_j'(x) + q(x)\varphi_j(x)]\varphi_i(x)dx = \\ &= \int_a^b \varphi_j''(x)\varphi_i(x)dx + \int_a^b p(x)\varphi_j'(x)\varphi_i(x)dx + \int_a^b q(x)\varphi_j(x)\varphi_i(x)dx. \end{aligned}$$

Первый из интегралов в этом выражении преобразуем «по частям»:

$$\int_a^b \varphi_j''(x) \varphi_i(x) dx = \varphi_i(x) \varphi_j'(x) \Big|_a^b - \int_a^b \varphi_j'(x) \varphi_i'(x) dx,$$

что приводит к следующей формуле для вычисления коэффициентов при неизвестных $c_1, c_2, \dots, c_n$ системы (10.59):

$$\begin{aligned} a_{ij} = & \varphi_i(b) \varphi_j'(b) - \varphi_i(a) \varphi_j'(a) - \int_a^b \varphi_j'(x) \varphi_i'(x) dx + \\ & + \int_a^b p(x) \varphi_j'(x) \varphi_i(x) dx + \int_a^b q(x) \varphi_j(x) \varphi_i(x) dx. \end{aligned} \quad (10.63)$$

Выполненное преобразование интеграла, во-первых, активизировало роль краевых условий при построении приближенного решения $y_n(x)$ в форме (10.40) (напомним, что функции $\varphi_i(x)$ при $i = 1, 2, \dots, n$ должны удовлетворять однородным краевым условиям (10.41)), во-вторых, позволило ослабить требования к гладкости базисных функций $\varphi_i(x)$, поскольку теперь, как видно из (10.63), у них могут быть даже разрывы производных (скачки). Подмеченный факт говорит о том, что методом Галёркина при подходящем выборе базисных функций можно находить решения краевых задач, определенные в каком-либо обобщенном смысле.

Пример 10.2. Снова обратимся к конкретной краевой задаче (10.55), к решению которой $y(x) = \frac{1}{x^2}$ в примере 10.1 было найдено квадратичное приближение методом коллокации.

Используя такое же представление

$$\tilde{y}_1(x) = \varphi_0(x) + \tilde{c}_1 \varphi_1(x) = \frac{13}{8} - \frac{5}{8}x + \tilde{c}_1 \left[(x-1)^2 - \frac{5}{4}(x-1) \right], \quad (10.64)$$

будем теперь искать его коэффициент $\tilde{c}_1$ методом Галёркина. Приведя данное уравнение в соответствие с видом (10.2), согласно которому имеем

$$p(x) = x^2, \quad q(x) = -x, \quad f(x) = \frac{6}{x^4} - \frac{3}{x},$$

по формуле (10.63) находим

$$a_{11} = \left(1 - \frac{5}{4}\right)\left(2 - \frac{5}{4}\right) - \int_1^2 \left[2(x-1) - \frac{5}{4}\right]^2 dx + \\ + \int_1^2 x^2 \left[2(x-1) - \frac{5}{4}\right] \cdot \left[(x-1)^2 - \frac{5}{4}(x-1)\right] dx - \\ - \int_1^2 x \left[(x-1)^2 - \frac{5}{4}(x-1)\right]^2 dx \approx -0.770,$$

а по формуле (10.62)

$$d_1 = \int_1^2 \left(\frac{6}{x^4} - \frac{3}{x} + \frac{13}{8}x\right) \left[(x-1)^2 - \frac{5}{4}(x-1)\right] dx \approx -0.548.$$

Следовательно, в (10.64)

$$\tilde{c}_1 = \frac{d_1}{a_{11}} \approx \frac{0.548}{0.770} \approx 0.712,$$

т.е. галёркинское квадратичное приближение к точному решению

$$y(x) = \frac{1}{x^2} \text{ есть}$$

$$\tilde{y}_1(x) \approx 1 - \frac{5}{8}(x-1) + 0.712(x-1)\left(x - \frac{9}{4}\right).$$

В точке $x = \frac{3}{2}$ таким способом получаем несколько более точное приближение $\tilde{y}_1\left(\frac{3}{2}\right) \approx 0.42$ к $y\left(\frac{3}{2}\right)$, чем найденное ранее методом коллокации.

Графическое сравнение квадратичных приближений $y_1(x)$ и $\tilde{y}_1(x)$ к решению $y(x)$ краевой задачи (10.55), полученных методом коллокации (см. пример 10.1) и методом Галёркина, представлено рисунком 10.3.

Замечание 10.5. Возможно, чаще в литературе, освещающей метод Галёркина для обыкновенных дифференциальных уравнений, исходное линейное уравнение второго порядка берут не в принятом здесь за основу виде (10.2), а в виде

$$-(k(x)y')' + q(x)y = f(x), \quad (10.65)$$

имеющем прямую физическую интерпретацию и называемом в связи с этим *одномерным стационарным уравнением теплопроводности* или *одномерным уравнением диффузии* (см. [2, 3, 39, 50, 55, 73, 86 и др.]).

При такой записи уравнения несколько упрощается выражение для подсчета коэффициентов a_{ij} системы (10.59) и, главное, становится более естественным анализ ситуации, когда коэффициенты в (10.65) могут оказаться разрывными.

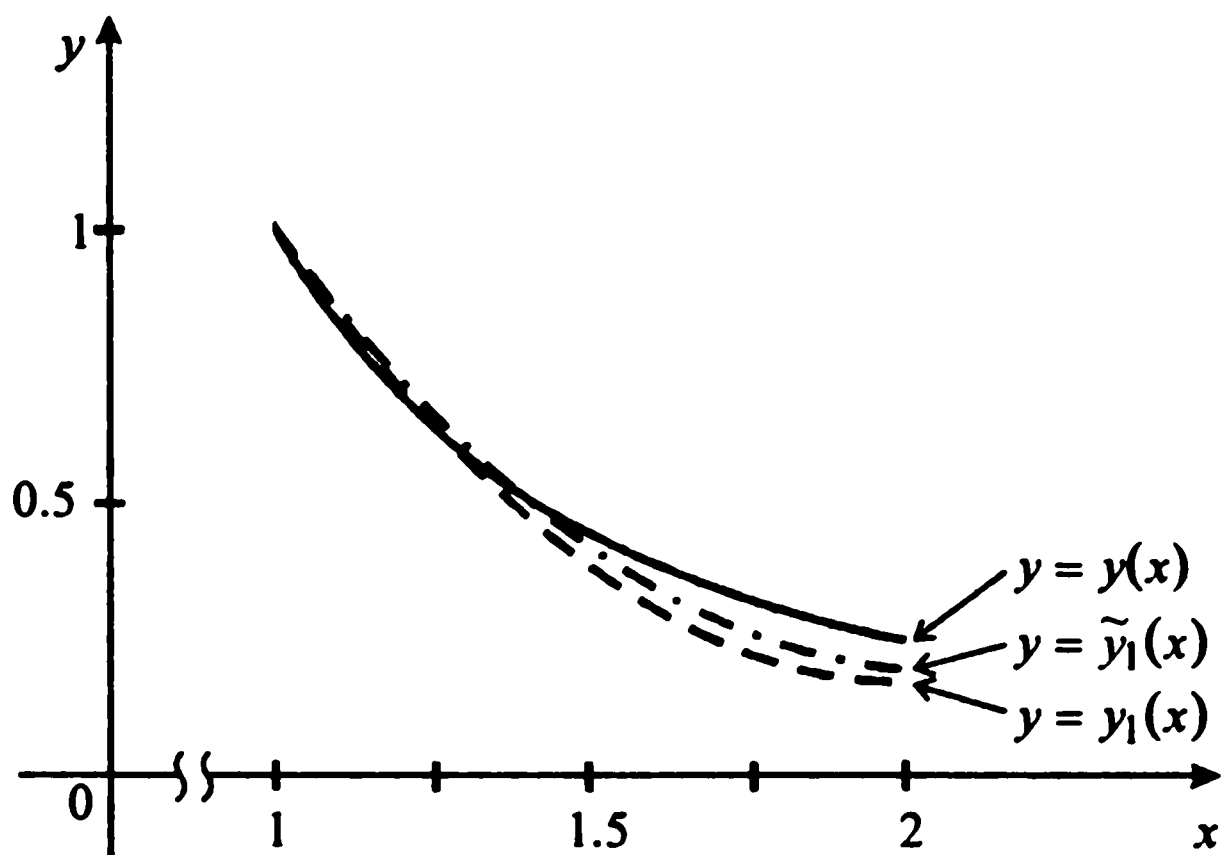


Рис. 10.3. Графики точного решения $y(x)$ краевой задачи (10.55) и его квадратичных приближений методами Галёркина ($\tilde{y}_1(x)$) и коллокации ($y_1(x)$)

Замечание 10.6. К той же системе (10.59) с коэффициентами (10.60) приходят и в *методе Рунца*, опирающемся на вариационный принцип приближенного решения краевой задачи (10.2)–(10.4). Сведения об этом методе, а также другие подходы и варианты метода Галёркина можно почерпнуть в книгах [12, 42, 53, 63, 64, 86 и др].

§ 10.6. МЕТОД КОНЕЧНЫХ ЭЛЕМЕНТОВ

При всех достоинствах метод Галёркина, рассмотренный в предыдущем параграфе, обладает тем существенным на современном этапе формализации прикладных исследований недостатком, что в «чистом» виде он малопригоден для автоматизированных компьютерных вычислений. Однако при подходящем выборе системы простых базисных функций φ_i в представлении приближенного решения $y_n(x)$, связанных с определенной на отрезке $[a, b]$ системой точек (сеткой), метод Галёркина трансформируется в сугубо численный процесс получения каркаса приближенного решения на заданной сетке, причем технология

построения этого каркаса в конечном итоге оказывается близкой к той, которая присуща методу конечных разностей (см. § 10.3). Отсюда — название такого численного процесса **проекционно-разностный** или **проекционно-сеточный метод**. Другое его более раннее и более употребительное название **метод конечных элементов** (МКЭ или FEM от англ. *finite element method*) связано с другими идеями, корни которых следует искать в строительной механике и теории упругости [72]. Преимущества этого метода перед многими другими методами обнаруживаются, в основном, при решении двумерных и трехмерных задач математической физики, однако первые представления о МКЭ кажется целесообразным получить именно здесь, развивая метод Галёркина для краевой задачи (10.2)–(10.4).

Введем на отрезке $[a, b]$ равномерную сетку с шагом $h = \frac{b-a}{n+1}$, состоящую из n внутренних точек (узлов) $x_i = a + ih$ ($i = 1, 2, \dots, n$) и двух крайних — $x_0 = a$, $x_{n+1} = b$.

Будем искать приближенное решение $y_n(x)$ данной краевой задачи (10.2)–(10.4) в виде линейной комбинации простых однотипных функций $y_i(x)$, на роль которых возьмем так называемые **финитные функции**, определяемые равенством

$$\varphi(t) = \begin{cases} 1 - |t|, & |t| \leq 1, \\ 0, & |t| > 1. \end{cases} \quad (10.66)$$

Считая переменную x принадлежащей отрезку $[a, b]$ и полагая в (10.66) $t = \frac{x - x_i}{h}$, видим, что

$$|t| < 1 \Leftrightarrow x_i - h < x < x_i + h,$$

т.е. функция $\varphi(x)$ отлична от нуля лишь на интервале (x_{i-1}, x_{i+1}) с центром в точке x_i . Таким образом, разные узлы x_i рассматриваемой сетки определяют разные функции $\varphi_i(x)$, которые, очевидно, согласно (10.66), можно задать равенствами

$$\varphi_i = \begin{cases} 1 + \frac{x - x_i}{h} = \frac{x - x_{i-1}}{h}, & \text{если } x \in [x_{i-1}, x_i], \\ 1 - \frac{x - x_i}{h} = -\frac{x - x_{i+1}}{h}, & \text{если } x \in [x_i, x_{i+1}], \\ 0, & \text{если } x \notin [x_{i-1}, x_{i+1}]. \end{cases} \quad (10.67)$$

График одной такой функции φ_i приведен на рис. 10.4.

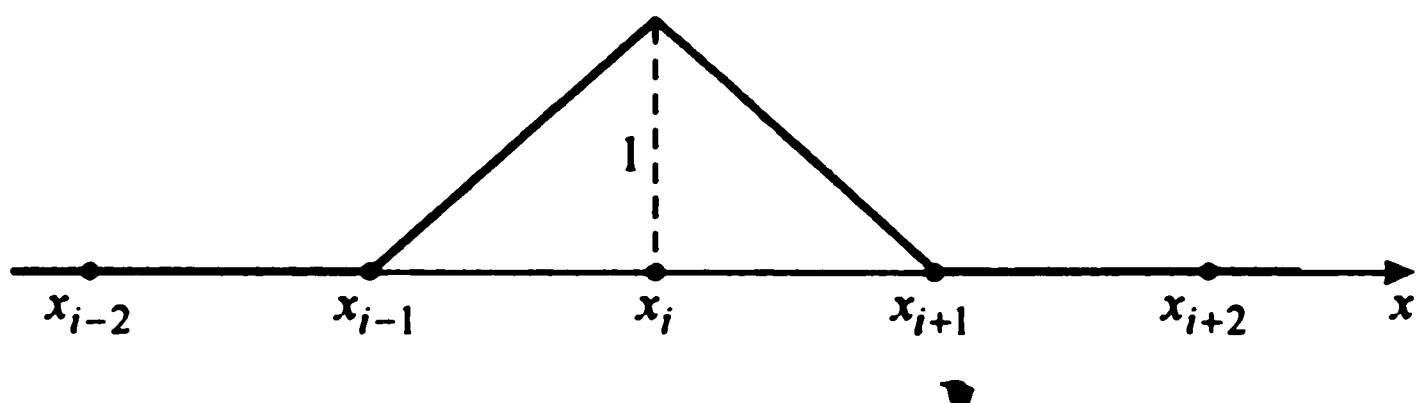


Рис. 10.4. График финитной функции φ_i

Сравнение рис. 10.4 с рис. 4.7 и формул (10.67) с формулами (4.62) показывает, что введенная здесь финитная функция, умноженная на постоянную $\frac{1}{h}$, совпадает с определенным в гл. 4 линейным B -сплайном.

Относительно совокупности финитных функций $\{\varphi_i\}$, задаваемых на отрезке $[a, b]$ формулами (10.67), известно, что они *линейно независимы* (более того, ортогональны в специальной энергетической норме [3]) и образуют полную систему^{*)} в пространстве $L_2[a, b]$. Это дает основание для их законного применения в качестве базисных функций метода Галёркина.

Для дальнейшей конкретизации описанного в общих чертах МКЭ нужно уточнить вид системы линейных алгебраических уравнений (10.59) относительно коэффициентов c_i , с которыми должны суммироваться выбранные базисные функции при получении приближенного решения, и из (10.60) вывести формулы коэффициентов этой системы. С этой целью продифференцируем (10.67), получив при этом

$$\varphi'_i = \begin{cases} \frac{1}{h} & \text{при } x \in (x_{i-1}, x_i), \\ -\frac{1}{h} & \text{при } x \in (x_i, x_{i+1}), \\ 0 & \text{при } x \notin (x_{i-1}, x_{i+1}), \end{cases} \quad (10.68)$$

и изобразим параллельно всю систему определенных на отрезке $[a, b]$ функций φ_i и их производных φ'_i (рис. 10.5).

^{*)} Говоря о полноте, имеем в виду процесс $h \rightarrow 0$.

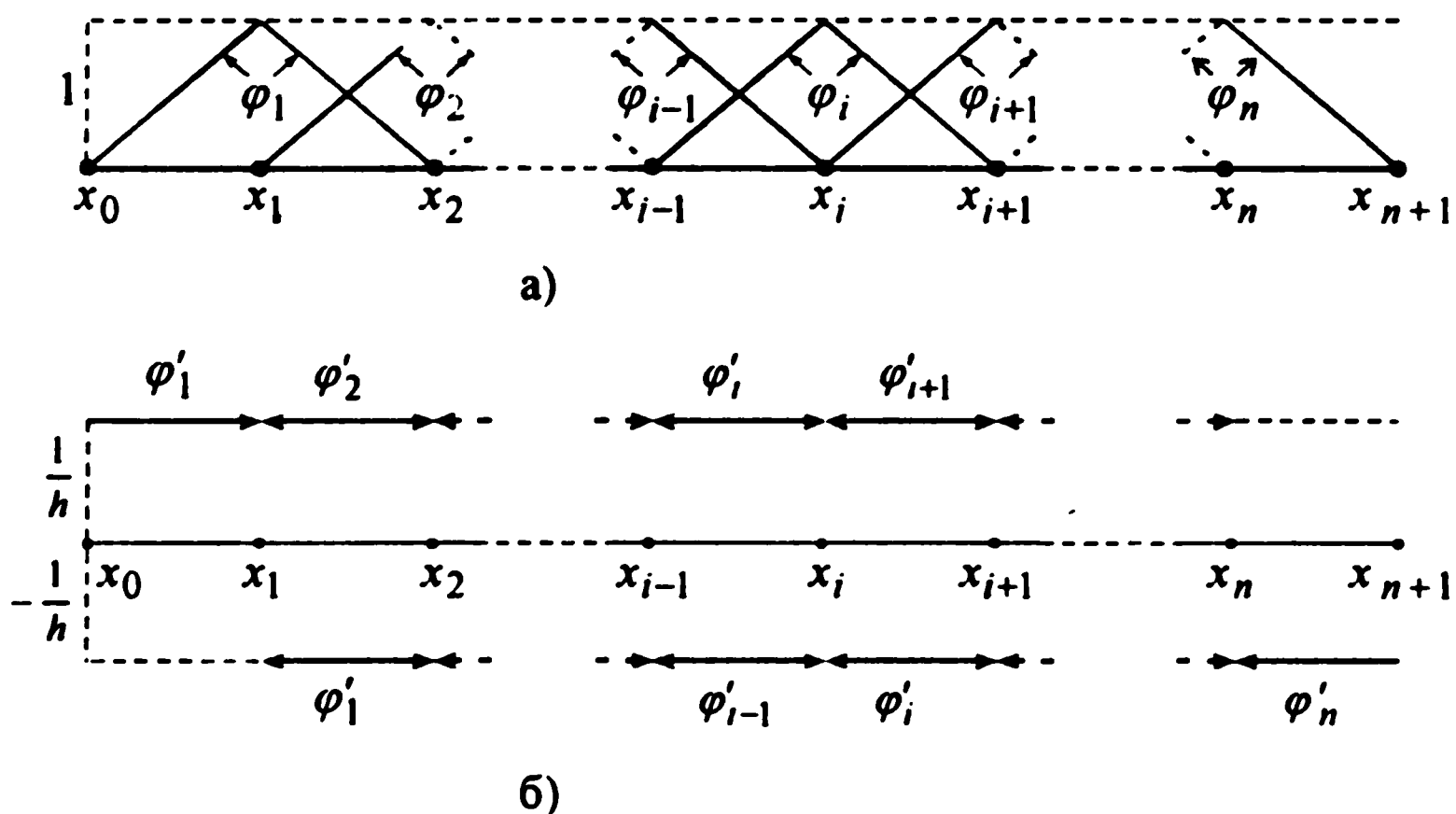


Рис. 10.5. Система финитных функций $\varphi_1, \varphi_2, \dots, \varphi_n$ (а) и их производных (б))

Из совместного рассмотрения графиков φ_i и φ'_i , наряду с уже отмеченным фактом, что i -я базисная функция и ее производная имеют ненулевые значения только на двух смежных элементарных промежутках (x_{i-1}, x_i) и (x_i, x_{i+1}) , видим также, что на одном элементарном промежутке ненулевыми являются две «соседние» базисные функции и их производные: на (x_{i-1}, x_i) отличны от нуля $\varphi_{i-1}, \varphi_i, \varphi'_{i-1}, \varphi'_i$, на (x_i, x_{i+1}) — $\varphi_i, \varphi_{i+1}, \varphi'_i, \varphi'_{i+1}$, и т.д. Это позволяет, применив свойство аддитивности по промежутку интегрирования к интегралам формул (10.62), (10.63) метода Галёркина, упростить их вычисление, отбросив заведомо нулевые слагаемые, и увидеть определенную структуру матрицы линейной системы (10.59), что облегчает ее решение.

Итак, пусть ищется приближенное решение наиболее простой краевой задачи для дифференциального уравнения (10.2), а именно, с однородными краевыми условиями первого рода

$$y(a) = 0, \quad y(b) = 0. \quad (10.69)$$

Для получения ее приближенного решения в виде

$$y_n(x) = \sum_{i=1}^n c_i \varphi_i(x) \quad (10.70)$$

с кусочно-линейными базисными функциями $\varphi_i = \varphi_i(x)$, определенными в (10.67), для подсчета коэффициентов c_i , согласно

методу Галёркина, нужно составить линейную алгебраическую систему (10.59). Ее правые части в таком случае суть

$$\begin{aligned}
 d_i &= \int_a^b f(x) \varphi_i(x) dx = \sum_{k=0}^n \int_{x_k}^{x_{k+1}} f(x) \varphi_i(x) dx = \\
 &= \int_{x_{i-1}}^{x_i} f(x) \frac{x - x_{i-1}}{h} dx + \int_{x_i}^{x_{i+1}} f(x) \left(-\frac{x - x_{i+1}}{h} \right) dx = \\
 &= \frac{1}{h} \left[\int_{x_{i-1}}^{x_i} f(x) (x - x_{i-1}) dx - \int_{x_i}^{x_{i+1}} f(x) (x - x_{i+1}) dx \right]. \quad (10.71)
 \end{aligned}$$

Так как при краевых условиях (10.69) используются n базисных функций с φ_1 по φ_n , и все они в точках a и b равны нулю (см. рис. 10.5), то формула (10.63) для вычисления элементов a_{ij} матрицы галёркинской СЛАУ (10.59) здесь имеет вид

$$\begin{aligned}
 a_{ij} &= - \int_a^b \varphi'_j(x) \varphi'_i(x) dx + \int_a^b p(x) \varphi'_j(x) \varphi_i(x) dx + \int_a^b q(x) \varphi_j(x) \varphi_i(x) dx = \\
 &= \sum_{k=0}^n \int_{x_k}^{x_{k+1}} \left[-\varphi'_j(x) \varphi'_i(x) + p(x) \varphi'_j(x) \varphi_i(x) + q(x) \varphi_j(x) \varphi_i(x) \right] dx. \quad (10.72)
 \end{aligned}$$

В силу отмеченного выше неравенства нулю на элементарном промежутке лишь соседних по индексу финитных функций и их производных, можно считать отличными от нуля фигурирующие в выражении a_{ij} произведения $\varphi'_j \varphi'_i$, $\varphi'_j \varphi_i$, $\varphi_j \varphi_i$, только в тех случаях, когда $i - 1 \leq j \leq i + 1$. А это означает, что

$$a_{ij} = 0 \quad \text{при} \quad |i - j| > 1, \quad (10.73)$$

т.е. матрица $A = (a_{ij})$ системы (10.59) имеет ленточную структуру, точнее, является классической трехдиагональной матрицей. Это позволяет применять для ее решения высокоэффективный метод прогонки, как и в методе конечных разностей.

Конкретизируем формулы для вычисления ненулевых элементов матрицы A . Полагая в (10.72) $j = i$, с помощью выражений (10.67), (10.68) получаем формулу для вычисления

диагональных элементов:

$$\begin{aligned}
 a_{ii} &= \int_{x_{i-1}}^{x_i} \left[-\frac{1}{h^2} + p(x) \frac{1}{h} \cdot \frac{x - x_{i-1}}{h} + q(x) \left(\frac{x - x_{i-1}}{h} \right)^2 \right] dx + \\
 &\quad + \int_{x_i}^{x_{i+1}} \left[-\left(-\frac{1}{h} \right)^2 + p(x) \frac{1}{h} \cdot \frac{x - x_{i+1}}{h} + q(x) \left(\frac{x - x_{i+1}}{h} \right)^2 \right] dx = \\
 &= -\frac{2}{h} + \frac{1}{h^2} \left[\int_{x_{i-1}}^{x_i} p(x)(x - x_{i-1}) dx + \int_{x_{i-1}}^{x_i} q(x)(x - x_{i-1})^2 dx + \right. \\
 &\quad \left. + \int_{x_i}^{x_{i+1}} p(x)(x - x_{i+1}) dx + \int_{x_i}^{x_{i+1}} q(x)(x - x_{i+1})^2 dx \right]. \quad (10.74)
 \end{aligned}$$

При $j = i + 1$ из (10.72) находим выражения элементов правой побочной диагонали матрицы **A** системы (10.59):

$$\begin{aligned}
 a_{i,i+1} &= \int_{x_i}^{x_{i+1}} \left[\frac{1}{h^2} + p(x) \frac{1}{h} \cdot \left(-\frac{x - x_{i+1}}{h} \right) + q(x) \left(-\frac{x - x_{i+1}}{h} \right) \cdot \frac{x - x_i}{h} \right] dx = \\
 &= \frac{1}{h} - \frac{1}{h^2} \left[\int_{x_i}^{x_{i+1}} p(x)(x - x_{i+1}) dx + \int_{x_i}^{x_{i+1}} q(x)(x - x_i)(x - x_{i+1}) dx \right], \quad (10.75)
 \end{aligned}$$

а при $j = i - 1$ — левой:

$$\begin{aligned}
 a_{i,i-1} &= \int_{x_{i-1}}^{x_i} \left[\frac{1}{h^2} + p(x) \left(-\frac{1}{h} \right) \frac{x - x_{i-1}}{h} + q(x) \left(-\frac{x - x_i}{h} \right) \cdot \frac{x - x_{i-1}}{h} \right] dx = \\
 &= \frac{1}{h} - \frac{1}{h^2} \left[\int_{x_{i-1}}^{x_i} p(x)(x - x_{i-1}) dx + \int_{x_{i-1}}^{x_i} q(x)(x - x_{i-1})(x - x_i) dx \right]. \quad (10.76)
 \end{aligned}$$

Таким образом, можно считать замысел МКЭ принципиально воплощенным, поскольку формулы (10.71), (10.73)–(10.76) полностью задают алгебраическую систему (10.59) для получения коэффициентов $c_1, \dots, c_n$ приближенного решения (10.70)

краевой задачи (10.2), (10.69). Сделаем лишь несколько замечаний.

Во-первых, находящаяся описанным способом функция $y_n(x)$ представляет собой *кусочно-линейную аппроксимацию* точного решения $y(x)$ краевой задачи (10.2), (10.69), а совокупность значений коэффициентов $c_1, \dots, c_n$ играет роль *каркаса* приближенного решения $y_n(x)$ на сетке $x_1, \dots, x_n$, что хорошо видно в результате подстановки $x = x_j$ в (10.70):

$$y_n(x_j) = \sum_{i=1}^n c_i \varphi_i(x_j) = c_j, \text{ в силу } \varphi_i(x_j) = \begin{cases} 1, & \text{если } j = i, \\ 0, & \text{если } j \neq i. \end{cases}$$

Во-вторых, при вычислении ненулевых элементов матрицы $A = (a_{ij})$ линейной системы (10.59) следует учитывать, что в формулах (10.74)–(10.76) имеются одинаковые интегралы, что хотя бы незначительно сокращает затраты при численном интегрировании.

В-третьих, в данном случае галёркинскую систему алгебраических уравнений (10.59), учитывая (10.73), удобно представить как краевую задачу для трехточечного разностного уравнения второго порядка, т.е. в виде

$$\begin{cases} a_{11}c_1 + a_{12}c_2 = d_1, \\ a_{i,i-1}c_{i-1} + a_{ii}c_i + a_{i,i+1}c_{i+1} = d_i \quad (i = 2, \dots, n-1), \\ a_{n,n-1}c_{n-1} + a_{nn}c_n = d_n, \end{cases} \quad (10.77)$$

свидетельствующем о готовности системы к применению метода прогонки [18].

В четвертых, при неоднородных краевых условиях первого рода

$$y(a) = A, \quad y(b) = B \quad (10.78)$$

можно воспользоваться описанным в § 10.4 приемом сведения задачи (10.2), (10.78) к задаче

$$L[u] = F(x),$$

где

$$F(x) = f(x) - p(x)v'(x) - q(x)v(x), \quad v(x) = A + \frac{B - A}{b - a}(x - a),$$

с однородными условиями

$$u(a) = 0, \quad u(b) = 0.$$

Найдя МКЭ ее приближенное решение

$$u_n(x) = \sum_{i=1}^n c_i \varphi_i(x),$$

получаем

$$y(x) \approx y_n(x) := u_n(x) + v(x).$$

Пример 10.3. Рассмотрим фигурирующее в примерах 10.1, 10.2 дифференциальное уравнение

$$y'' + x^2 y' - xy = \frac{6}{x^4} - \frac{3}{x}, \quad x \in [1, 2] \quad (10.79)$$

при краевых условиях первого рода

$$y(1) = 1, \quad y(2) = 0.25. \quad (10.80)$$

К этой задаче, имеющей то же точное решение $y(x) = \frac{1}{x^2}$, что и в предыдущей задаче (10.55) для этого уравнения, применим метод конечных элементов с двумя базисными функциями φ_1 и φ_2 .

Сначала выполним преобразование данной задачи к задаче с однородными условиями

$$u'' + x^2 u' - xu = F(x), \quad x \in [1, 2], \quad u(1) = 0, \quad u(2) = 0, \quad (10.81)$$

для чего делаем замену $y = u + v$, где

$$v = 1 + \frac{0.25 - 1}{2 - 1}(x - 1) = \frac{7}{4} - \frac{3}{4}x,$$

и, следовательно,

$$F(x) = \frac{6}{x^4} - \frac{3}{x} - x^2 v' + xv = \frac{6}{x^4} - \frac{3}{x} + \frac{7}{4}x.$$

Введя на отрезке $[1, 2]$ равномерную сетку

$$x_0 = 1, \quad x_1 = \frac{4}{3}, \quad x_2 = \frac{5}{3}, \quad x_3 = 2 \quad (10.82)$$

с шагом $h = \frac{1}{3}$, записываем выражение приближенного решения задачи (10.81)

$$u_2(x) = c_1 \varphi_1(x) + c_2 \varphi_2(x),$$

где $\varphi_1(x)$ и $\varphi_2(x)$ — соответствующие сетке (10.82) функции-«крышки» (10.67). Для получения коэффициентов c_1, c_2 их линейной комбинации в

соответствии с (10.77) (при $n = 2$) составляем линейную алгебраическую систему^{*)}

$$\begin{cases} a_{11}c_1 + a_{12}c_2 = d_1, \\ a_{21}c_1 + a_{22}c_2 = d_2. \end{cases} \quad (10.83)$$

За числовыми данными этой системы обращаемся к формулам (10.71), (10.74)–(10.76), согласно которым имеем:

$$a_{11} = -6 + 9 \cdot \left[\int_1^{4/3} x^2(x-1)dx + \int_1^{4/3} (-x)(x-1)^2 dx + \right. \\ \left. + \int_{4/3}^{5/3} x^2\left(x - \frac{5}{3}\right)dx + \int_{4/3}^{5/3} (-x)\left(x - \frac{5}{3}\right)^2 dx \right] \approx -6.5926,$$

$$a_{22} = -6 + 9 \cdot \left[\int_{4/3}^{5/3} x^2\left(x - \frac{4}{3}\right)dx + \int_{4/3}^{5/3} (-x)\left(x - \frac{4}{3}\right)^2 dx + \right. \\ \left. + \int_{5/3}^2 x^2(x-2)dx + \int_{5/3}^2 (-x)(x-2)^2 dx \right] \approx -6.7407,$$

$$a_{12} = 3 - 9 \cdot \left[\int_{4/3}^{5/3} x^2\left(x - \frac{5}{3}\right)dx + \int_{4/3}^{5/3} (-x)\left(x - \frac{4}{3}\right)\left(x - \frac{5}{3}\right)dx \right] \approx 3.9630,$$

$$a_{21} = 3 - 9 \cdot \left[\int_{4/3}^{5/3} x^2\left(x - \frac{4}{3}\right)dx + \int_{4/3}^{5/3} (-x)\left(x - \frac{4}{3}\right)\left(x - \frac{5}{3}\right)dx \right] \approx 1.7037,$$

$$d_1 = 3 \int_1^{4/3} \left(\frac{6}{x^4} - \frac{3}{x} + \frac{7}{4}x \right) (x-1)dx - 3 \int_{4/3}^{5/3} \left(\frac{6}{x^4} - \frac{3}{x} + \frac{7}{4}x \right) \left(x - \frac{5}{3} \right) dx \approx 0.7256,$$

$$d_2 = 3 \int_{4/3}^{5/3} \left(\frac{6}{x^4} - \frac{3}{x} + \frac{7}{4}x \right) \left(x - \frac{4}{3} \right) dx - 3 \int_{5/3}^2 \left(\frac{6}{x^4} - \frac{3}{x} + \frac{7}{4}x \right) (x-2)dx \approx 0.6446.$$

^{*)} При таком малом числе базисных функций, разумеется, ни о какой ленточной структуре матрицы этой системы не может быть и речи; здесь даже нет ни одного «полного» трехточечного уравнения.

В результате подстановки этих чисел в систему (10.83) находим

$$c_1 \approx -0.1976, \quad c_2 \approx -0.1456.$$

Таким образом, приближенное решение $y_2(x)$ данной задачи (10.79), (10.80) есть

$$y_2(x) = v(x) + u_2(x) \approx 1.75 - 0.75x - 0.1976\varphi_1(x) - 0.1456\varphi_2(x),$$

где

$$\begin{aligned} \varphi_1(x) &= \begin{cases} 3(x-1), & \text{при } x \in \left[1, \frac{4}{3}\right], \\ -3\left(x-\frac{5}{3}\right), & \text{при } x \in \left[\frac{4}{3}, \frac{5}{3}\right], \end{cases} \\ \varphi_2(x) &= \begin{cases} 3\left(x-\frac{4}{3}\right), & \text{при } x \in \left[\frac{4}{3}, \frac{5}{3}\right], \\ -3(x-2), & \text{при } x \in \left[\frac{5}{3}, 2\right]. \end{cases} \end{aligned} \tag{10.84}$$

Последнее можно преобразовать в канонический вид кусочно-линейной на отрезке $[1, 2]$ функции

$$y_2(x) \approx \begin{cases} 2.3428 - 1.3428x, & \text{если } x \in \left[1, \frac{4}{3}\right], \\ 1.3444 - 0.5940x, & \text{если } x \in \left[\frac{4}{3}, \frac{5}{3}\right], \\ 0.8764 - 0.3132x, & \text{если } x \in \left[\frac{5}{3}, 2\right]. \end{cases} \tag{10.85}$$

Чтобы получить представление о точности найденного приближенного решения $y_2(x)$, сравним его значения с точными во внутренних узлах сетки и в середине заданного промежутка, что отразим следующей таблицей:

x	$\frac{4}{3}$	$\frac{3}{2}$	$\frac{5}{3}$
$y_2(x)$	0.5524	0.4534	0.3544
$y(x)$	0.5625	0.4444	0.36

Изображения базисных функций φ_1, φ_2 (штрих-пунктирные линии), посредством которых методом конечных элементов получено кусочно-линейное приближение $y_2(x)$ к точному решению $y(x)$ данной в примере

краевой задачи, и самих этих функций $y_2(x)$ и $y(x)$ (штриховая и сплошная линии соответственно) представлены на рис. 10.6.

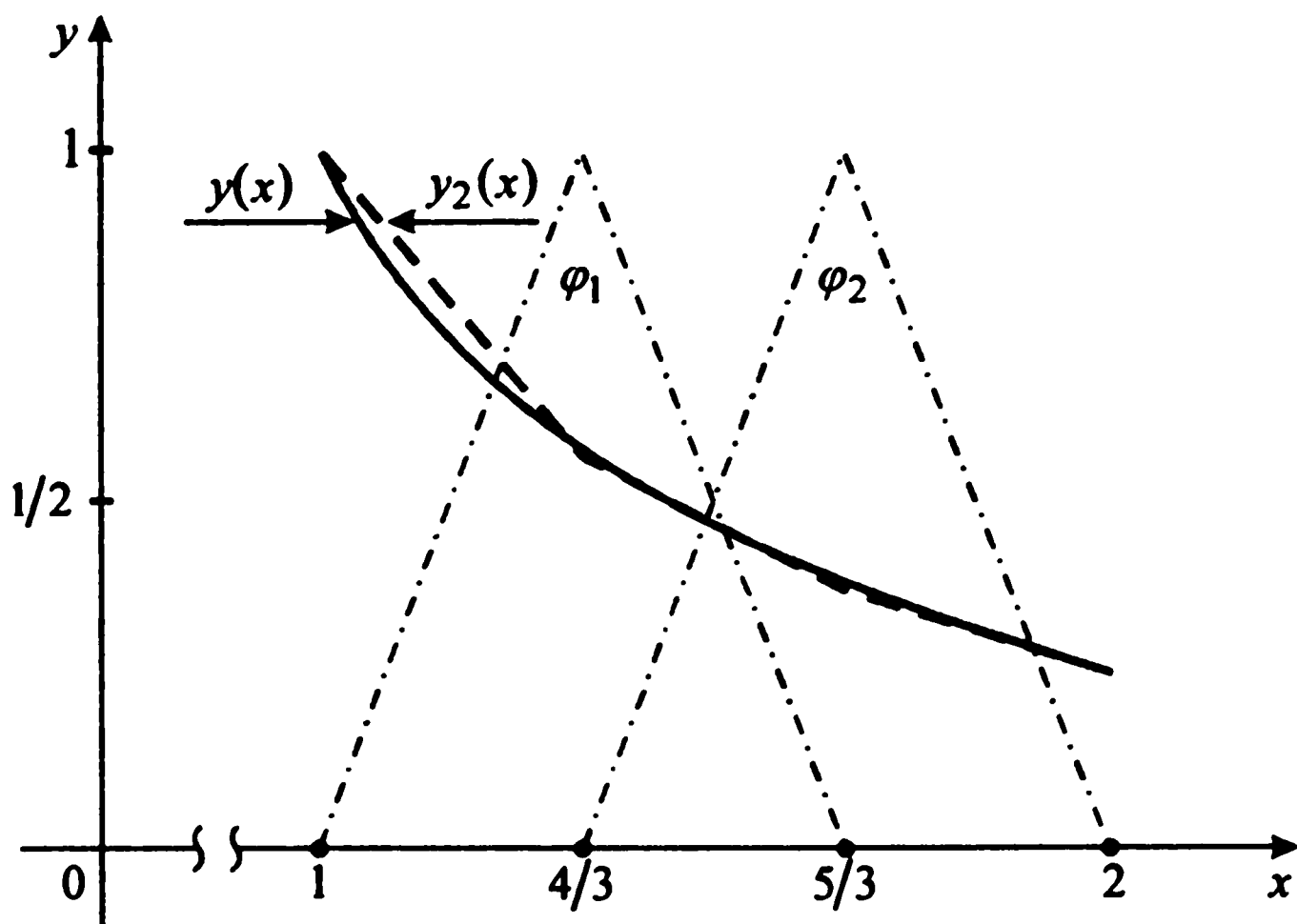


Рис. 10.6. Графики точного решения задачи (10.79)–(10.80), приближения к нему МКЭ (10.85) и базисных функций (10.84)

Увеличив количество узлов и базисных функций на единицу, приходим к СЛАУ

$$\begin{cases} -8.4167c_1 + 4.8333c_2 & = 0.5992, \\ 2.9375c_1 - 8.5000c_2 + 5.1875c_3 & = 0.4578, \\ & 2.5417c_2 - 8.5833c_3 = 0.5020. \end{cases}$$

Найдя из нее значения c_1 , c_2 , c_3 , получаем приближенное решение

$$y_3(x) \approx 1.75 - 0.75x - 0.1725\varphi_1(x) - 0.1806\varphi_2(x) - 0.1110\varphi_3(x)$$

с базисными функциями φ_1 , φ_2 , φ_3 , легко конкретизируемыми посредством формулы (10.67) с учетом того, что узлами сетки в этом случае служат точки

$$x_0 = 1, \quad x_1 = 1.25, \quad x_2 = 1.5, \quad x_3 = 1.75, \quad x_4 = 2.$$

Приближение $y_3(x)$ уже существенно точнее построенного выше приближения $y_2(x)$: его каркас на данной сетке совпадает с каркасом точного решения $y(x)$ задачи (10.79), (10.80) во всех контролируемых здесь четырех десятичных знаках.

Замечание 10.7. Не касаясь особо вопросов, связанных со сходимостью и с точностью метода конечных элементов (доказано, что при определенных условиях можно считать $\|y(x) - y_n(x)\| = O(h^2)$), представим себе возможность его развития за счет использования B -сплайнов порядков, больше первого. Можно ожидать, что при таком повышении степени базисных функций будет повышаться точность приближенного решения с одновременным усложнением формул для подсчета коэффициентов линейной алгебраической системы, сохраняющей ленточную структуру, но с увеличивающейся шириной ленты из ненулевых элементов. Имеются также модификации МКЭ, приспособленные к эффективному решению краевых задач для дифференциальных уравнений с заведомо разрывными коэффициентами; первые представления об этом можно получить из учебного пособия [2].

Замечание 10.8. Декларируемая в начале параграфа равномерность сетки, обеспечивающая одинаковость *конечных элементов*, коими в одномерном случае являются элементарные промежутки $[x_{i-1}, x_i]$, нужна была лишь для некоторого упрощения формул. Отказавшись от симметричности задания функции $\varphi(t)$ в (10.66), без большого труда можно распространить все рассуждения и выкладки на произвольные сетки (с переменным шагом $h_i := x_i - x_{i-1}$).

УПРАЖНЕНИЯ

10.1. Составьте алгоритм решения методом пристрелки (с точностью $\varepsilon > 0$ удовлетворения правому краевому условию) краевой задачи:

а) $x^2 y'' \ln(x) - xy' + y = 0, \quad x \in [1, e], \quad y(1) = 0, \quad y(e) = e - 2;$

б) $y'' = \sqrt{y'}, \quad x \in [0, 2], \quad y(0) = 0, \quad y(2) = \frac{2}{3}.$

10.2. К краевой задаче

$$y'' - y = x, \quad x \in [0, 1], \quad 2y(0) - y'(0) = 1, \quad y(1) = 2$$

примените: а) метод редукции;
б) метод дифференциальной прогонки.

10.3. Запишите формулы, определяющие для краевой задачи (10.2)–(10.4) метод правой дифференциальной прогонки (см. замечание 10.2). Примените их к решению задачи

$$y'' + \frac{1}{x} y' - \frac{3}{x^2} y = -\frac{3}{4\sqrt{x}}, \quad x \in [1, 4], \quad y(1) = 1, \quad y(4) - 2y'(4) = 2.$$

10.4. Методом конечных разностей второго порядка составьте алгебраическую систему уравнений относительно значений решения краевой задачи

$$y'' + 2y' - 3xy = \frac{2-8x}{x^3}, \quad x \in [1, 2], \quad y(1) = 1, \quad y(2) = 0.5 \quad (10.86)$$

на сетке с шагом $h = 0.2$.

10.5. Дана краевая дифференциальная задача

$$y'' - y' \ln(x) - 2y = 1, \quad x \in [0.5, 1.5],$$

$$y(0.5) + y'(0.5) = 1, \quad y(1.5) - y'(1.5) = 0.$$

- 1) Аппроксимируйте ее разностной задачей с помощью МКР второго порядка на сетке с шагом $h = 0.125$. Подготовьте полученную СЛАУ к прогонке. Можно ли гарантировать устойчивость прогонки?
- 2) Примените противопотоковый метод на той же сетке.

10.6. Выполните конечноразностные аппроксимации краевой задачи

$$y'' + p(x)y' + q(x)y = f(x), \quad x \in [a, b], \quad y(a) = A, \quad y(b) = B$$

третьего и четвертого порядков точности. Какие можно отметить достоинства и недостатки построенных разностных схем? Опробуйте схему четвертого порядка на задаче (10.86).

10.7. Убедитесь в справедливости формул (10.51), (10.52) для подсчета параметров γ_i базисных функций (10.49), (10.50), опираясь на второе из однородных краевых условий (10.41).

10.8. Для краевой задачи (10.55) примера 10.1 (см. § 10.4), используя фигурирующие там функции φ_0 и φ_1 , методом коллокации найдите квадратичное приближение $\hat{y}_1(x)$, взяв за основу другой узел коллокации $\hat{x}_1 = \frac{7}{4}$. Сравните значения $\hat{y}_1\left(\frac{3}{2}\right)$, $\hat{y}_1\left(\frac{7}{4}\right)$ и $\hat{y}_1(2)$ с соответствующими значениями точного решения $y(x)$ и найденного в примере 10.1 коллокацией в узле $x_1 = \frac{3}{2}$ приближенного решения $y_1(x)$.

10.9. Полагая $h = 0.25$, примените к краевой задаче (10.86):

- а) метод коллокации со степенными базисными функциями;
- б) метод Галёркина с теми же базисными функциями.

Сравните результаты в узлах сетки.

10.10. Выведите формулы для вычисления коэффициентов СЛАУ (10.58), к решению которой сводится применение метода Галёркина для дифференциального уравнения (10.65) с краевыми условиями первого рода (10.7) (см. замечание 10.5). На их основе получите соответствующие формулы метода конечных элементов для этой задачи.

10.11. Обобщите расчетные формулы метода конечных элементов так, чтобы они были пригодны для решения задачи (10.2), (10.78) в случае неравномерной сетки (см. замечание 10.8).

10.12. Дана краевая задача

$$(x+3)^2 y'' + (2x+6)y' + 0.25y = \sqrt{x+3}, \quad x \in [0, 1],$$
$$y(0) = 0, \quad y(1) = 0.5.$$

А) Примените метод конечных элементов с одной, с двумя и с тремя базисными функциями на равномерной сетке. Сравните сеточные значения найденных приближений $y_1(x)$, $y_2(x)$, $y_3(x)$ и точного решения $y(x) = x/\sqrt{x+3}$.

Б) В случае единственного внутреннего узла (одной базисной функции) опробуйте результаты выполнения упр.10.11, принимая за этот узел поочередно точки $\frac{1}{3}$ и $\frac{2}{3}$ и сравнивая полученные приближения между собой и с $y_1(x)$ из п.А).

ГЛАВА 11 || ЧИСЛЕННОЕ РЕШЕНИЕ || ИНТЕГРАЛЬНЫХ УРАВНЕНИЙ

Даются некоторые общие понятия об интегральных уравнениях, в частности, о линейных уравнениях Фредгольма и Вольтерра. Описывается способ решения интегральных уравнений частного вида, известный как метод вырожденных ядер. Основное внимание уделяется численному решению уравнений методом конечных сумм, базирующемуся на использовании тех или иных квадратурных формул. Такой достаточно универсальный метод позволяет сводить данную задачу в функционально-интегральной постановке к решению линейной алгебраической системы относительно каркаса приближенного решения: с квадратной матрицей коэффициентов в случае уравнений Фредгольма и с треугольной — для уравнений Вольтерра. Изложение сопровождается числовыми примерами. Последний параграф главы посвящен квадратурно-итерационному методу вычисления каркаса резольвенты. Этот метод опирается на дискретизацию (на квадратурной основе) известной связи между резольвентой и ядром уравнения второго рода с последующим применением итерационного процесса Шульца.

§ 11.1. НЕКОТОРЫЕ ОБЩИЕ СВЕДЕНИЯ ОБ ИНТЕГРАЛЬНЫХ УРАВНЕНИЯХ^{*)}

Интегральным уравнением называют уравнение относительно неизвестной функции, содержащейся под знаком интеграла. С частными интегральными уравнениями мы уже встречались в главах 7, 8, преобразуя к интегральной форме задачу Коши для дифференциального уравнения (см. (7.3), (8.53)).

К интегральным уравнениям приводят многие задачи, возникающие и в самой математике, и в многочисленных ее приложениях. Исторически первой задачей, оформленной как интегральное уравнение

$$\int_0^z \frac{\varphi(\eta)}{\sqrt{z-\eta}} d\eta = f(z), \quad (11.1)$$

считается **задача Абеля** (1823 г), заключающаяся в определении вида кривой $x = \varphi(z)$, по которой в вертикальной плоскости Oxz под действием силы тяжести скатывается материальная точка так, чтобы начав свое движение без начальной скорости в точке

^{*)} Более подробно смотри в источниках [19, 36, 51, 59, 60, 68, 93 и др.]

кривой с аппликатой z , она достигала оси Ox за заданное время $T = f(z)$.

Другим примером может служить описание зависимостей между напряжениями σ и деформациями ε упруго-вязких материалов уравнениями

$$\varepsilon(t) = \frac{\sigma(t)}{E} + \frac{1}{E} \int_0^t K(t - \tau) \sigma(\tau) d\tau, \quad (11.2)$$

$$\sigma(t) = E\varepsilon(t) - E \int_0^t T(t - \tau) \varepsilon(\tau) d\tau, \quad (11.3)$$

где E — модуль упругости, $K(t - \tau)$ — функция влияния напряжения $\sigma(t)$ в момент τ на деформацию $\varepsilon(t)$ в момент t , $T(t - \tau)$ — аналогичная функция влияния деформации.

Основанием для составления уравнений обычно служат общие физические (в широком смысле) законы. Известные законы сохранения массы, импульса, энергии имеют интегральную формулировку и приводят к интегральным уравнениям. Интегральными уравнениями отражаются законы газовой динамики, электродинамики, экологии и т.д. Достоинством таких моделей служит то обстоятельство, что интегральные уравнения, в отличие от дифференциальных, не содержат производных искомой функции, а значит, не накладывают жестких ограничений на гладкость решения.

Достаточно общий вид интегральных уравнений содержится в записи

$$x(t) = \int_D K(t, s, x(s)) ds + f(t), \quad (11.4)$$

где D — некоторая область n -мерного пространства, x — неизвестная, а f — известная векторные функции; K — в общем случае нелинейная относительно x функция.

Ограничимся рассмотрением лишь одномерных уравнений, т.е. таких, в которых искомой неизвестной является скалярная функция одной переменной и интегрирование производится по отрезку. При этом будем иметь в виду только риманов интеграл, хотя часто для большей общности интегральные уравнения изучаются с привлечением интеграла Лебега. Более того, наше внимание не выйдет за рамки **линейных интегральных уравнений**, т.е. таких уравнений (11.4), в которых подынтегральная функция $K(t, s, x(s))$ представима в виде $Q(t, s)x(s)$.

Опишем основные типы линейных интегральных уравнений. По тому, постоянны ли обе границы интегрирования или одна из них может быть переменной, линейные интегральные

уравнения подразделяются на уравнения Фредгольма и уравнения Вольтерра соответственно^{*)}. Более простыми (с более хорошими свойствами) и более широко применяемыми являются **линейные интегральные уравнения второго рода**:

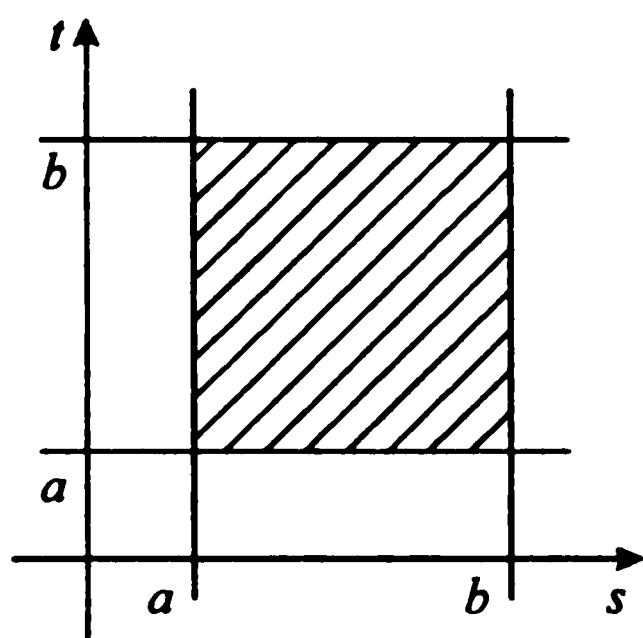
Фредгольма —

$$x(t) = \lambda \int_a^b Q(t, s)x(s)ds + f(t) \quad (11.5)$$

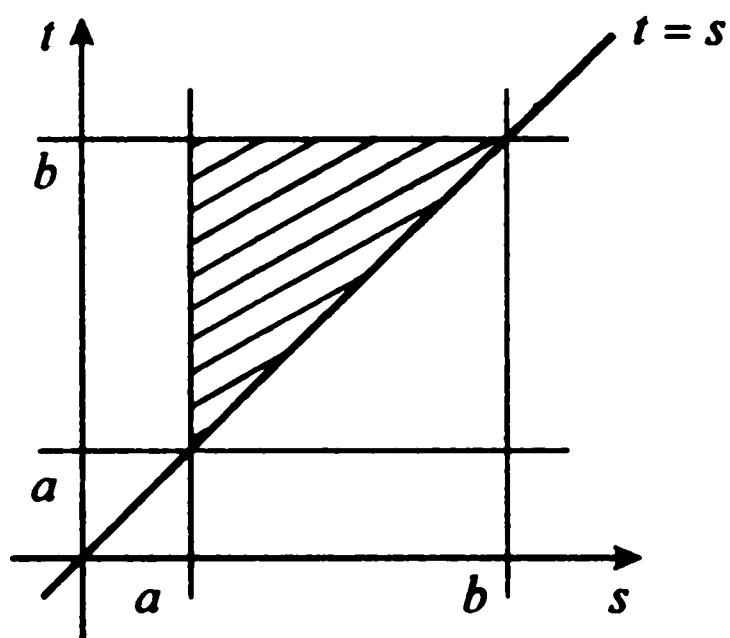
и Вольтерра —

$$x(t) = \lambda \int_a^t Q(t, s)x(s)ds + f(t). \quad (11.6)$$

Заданная функция $f(t)$ — свободный член — и неизвестная функция $x(t)$ — решение — в этих уравнениях зависят от переменной t , изменяющейся на отрезке $[a, b]$. Функция двух переменных $Q(t, s)$, называемая **ядром** интегрального уравнения, определяется на множестве точек квадрата $[a, b] \times [a, b]$ в случае интегрального уравнения Фредгольма (рис.11.1, а) и треугольника $a \leq s \leq t \leq b$ в случае уравнения Вольтерра (рис.11.1, б)).



а)



б)

Рис.11.1. Области задания ядер $Q(t, s)$ интегральных уравнений Фредгольма (а) и Вольтерра (б)

^{*)} Фредгольм Эрик Ивар (1866–1927) — шведский математик; Вольтерра Вито (1860–1940) — итальянский математик. Устоявшееся в течение многих лет русское написание «уравнение Вольтерра» все чаще сменяется написанием «уравнение Вольтерры». Отдадим дань традиции в ущерб правописанию.

Сразу заметим, что доопределив ядро $Q(t, s)$ уравнения Вольтерра нулем при $a \leq s \leq t \leq b$ (в нештрихованной части квадрата $[a, b] \times [a, b]$ на рис.11.1, б)), уравнение Вольтерра можно посчитать уравнением Фредгольма и применять к нему известные для уравнений Фредгольма результаты. Однако при таком подходе теряется специфика уравнений Вольтерра, и возникают сложности, которых можно избежать при отдельном изучении этих уравнений^{*)}.

Введение в линейное интегральное уравнение числового параметра λ (который можно отнести и к ядру) придает уравнению более общий вид и позволяет установить теоремы существования решений при тех или иных значениях λ , в то время как при нейтральном значении $\lambda = 1$ решения может не оказаться. Подобно тому, как это делается в теориях линейных систем алгебраических или линейных дифференциальных уравнений, существование и единственность решений линейных неоднородных уравнений (11.5), (11.6) изучается посредством изучения соответствующих им однородных уравнений, т.е. уравнений, получающихся из (11.5), (11.6) при $f(t) \equiv 0$.

Более сложно обстоит дело с существованием решений, их единственностью и непрерывной зависимостью решений от правой части для *интегральных уравнений первого рода*:

Фредгольма —

$$\int_a^b Q(t, s)x(s)ds = f(t) \quad (11.7)$$

и Вольтерра —

$$\int_a^t Q(t, s)x(s)ds = f(t). \quad (11.8)$$

Такие уравнения, характеризующиеся отсутствием отдельного слагаемого $x(t)$ (не связанного интегралом), имеют более ограниченную сферу применения и являются наиболее типичными представителями *некорректных задач*. Особенно это касается уравнений Фредгольма (11.7), в которых, кстати, иногда считают, что независимая переменная t изменяется не на промежутке интегрирования $[a, b]$, а на некотором другом отрезке $[c, d]$, где $c = a \pm \delta_1$, $d = b \pm \delta_2$, т.е. ядро $Q(t, s)$ задается на

^{*)} Выше уже отмечалась связь между задачами Коши для дифференциальных уравнений и интегральными уравнениями с переменными верхними границами, кои являются уравнениями Вольтерра. Аналогично, интегральные уравнения Фредгольма связывают с краевыми задачами для дифференциальных уравнений. Отсюда может проистекать понимание, например, того, насколько жестче стоит вопрос о разрешимости уравнения Фредгольма по сравнению с разрешимостью уравнения Вольтерра.

прямоугольнике $[c, d] \times [a, b]$.

Уравнения первого и второго рода можно объединить общей записью

$$h(t)x(t) = \lambda \int_D Q(t, s)x(s)ds + f(t). \quad (11.9)$$

При $h(t) \equiv 0$ формула (11.9) определяет уравнение первого рода, при $h(t) \equiv 1$ — второго рода; если же $h(t)$ обращается в нуль в некоторых точках промежутка интегрирования D , то (11.9) есть **уравнение третьего рода**, более редко встречающееся в приложениях и менее изученное.

Очень многие используемые на практике уравнения выделяются тем, что их ядра зависят только от разности $t - s$. Такие интегральные уравнения называют уравнениями с разностным ядром. К ним, очевидно, относятся приведенные в начале параграфа конкретные интегральные уравнения Вольтерра (11.1)–(11.3) (первое из них — уравнение первого рода, второе и третье — второго рода).

При непрерывных функциях $Q(t, s)$ и $f(t)$ достаточно просто устанавливается существование единственного непрерывного решения $x(t)$ уравнения Вольтерра второго рода (11.6) при любых значениях параметра λ ; отсюда — несущественность его введения в эти уравнения. Для уравнений Фредгольма второго рода (11.5) при тех же требованиях непрерывности $Q(t, s)$ и $f(t)$ существование единственного непрерывного решения можно установить, например, при условии, что

$$|\lambda| < \frac{1}{C(b-a)}, \quad \text{где} \quad C := \max_{t, s \in [a, b]} |Q(t, s)|. \quad (11.10)$$

При снижении требований к гладкости возможных решений условие (11.10) ослабляется. Например, в случае функций, интегрируемых с квадратом, в роли достаточного условия вместо (11.10) часто фигурирует неравенство

$$|\lambda| < \frac{1}{\sqrt{\int_a^b \int_a^b Q^2(t, s) dt ds}}.$$

Для некоторых уравнений второго рода с ядрами определенной структуры имеются формулы (или совокупности формул), позволяющие найти точное решение $x(t)$.

Пример 11.1. [52]. Рассмотрим уравнение

$$x(t) = \lambda \int_0^{\pi} \sin(t-s)x(s)ds + \cos(t),$$

являющееся линейным интегральным уравнением Фредгольма второго рода. Его разностное ядро $\sin(t-s)$ по формуле синуса разности можно преобразовать к вырожденному:

$$\sin(t-s) = \sin t \cos s - \cos t \sin s.$$

Сопоставляя это с (11.13), видим, что если данное уравнение имеет решение, то его следует искать в соответствующей (11.14) форме:

$$x(t) = \lambda z_1 \sin t - \lambda z_2 \cos t + \cos t. \quad (11.17)$$

По формулам (11.16) вычисляем:

$$\begin{aligned} c_{11} &:= \int_0^{\pi} \cos s \sin s ds = 0, & c_{12} &:= -\int_0^{\pi} \cos^2 s ds = -\frac{\pi}{2}, \\ c_{21} &:= \int_0^{\pi} \sin^2 s ds = \frac{\pi}{2}, & c_{22} &:= -\int_0^{\pi} \sin s \cos s ds = 0, \\ d_1 &:= \int_0^{\pi} \cos^2 s ds = \frac{\pi}{2}, & d_2 &:= \int_0^{\pi} \sin s \cos s ds = 0. \end{aligned}$$

Составив систему вида (11.15)

$$\begin{cases} z_1 + \frac{\lambda\pi}{2} z_2 = \frac{\pi}{2}, \\ -\frac{\lambda\pi}{2} z_1 + z_2 = 0 \end{cases}$$

и вычислив определитель Фредгольма

$$D(\lambda) = \begin{vmatrix} 1 & \frac{\lambda\pi}{2} \\ -\frac{\lambda\pi}{2} & 1 \end{vmatrix} = 1 + \frac{\lambda^2 \pi^2}{4},$$

убеждаемся, что он отличен от нуля при любых вещественных значениях λ и, значит, решение (11.17) существует и единственно. Найдя еще определители

$$D_1(\lambda) = \frac{\pi}{2}, \quad D_2(\lambda) = \frac{\lambda\pi^2}{4},$$

по формулам Крамера записываем значения коэффициентов

$$z_1 = \frac{2\pi}{4 + \lambda^2 \pi^2}, \quad z_2 = \frac{\lambda\pi^2}{4 + \lambda^2 \pi^2},$$

подстановка которых в выражение (11.17) приводит к решению

$$x(t) = \frac{4 \cos t + 2\lambda\pi \sin t}{4 + \lambda^2 \pi^2} \quad \forall \lambda \in \mathbb{R}.$$

Умение находить точное решение интегрального уравнения с вырожденным ядром порождает приближенный метод, в основе которого лежит замена одного уравнения другим, ядро которого вырожденно и в каком-то определенном смысле близко к данному. Такая замена ядра вырожденным может опираться на разные способы локальной аппроксимации функции двух переменных.

Например, уравнение

$$x(t) = \int_0^{0.5} t \sin(ts) x(s) ds + \cos 0.5t \quad (11.18)$$

с определенной степенью точности (которую доступно оценить) можно заменить уравнением с вырожденным ядром

$$x(t) = \int_0^{0.5} t^2 s x(s) ds + \cos 0.5t, \quad (11.19)$$

пользуясь тем, что по формуле Тейлора при малых ts имеет место приближенное равенство

$$\sin(ts) \approx ts + O((ts)^3). \quad (11.20)$$

Кроме вышеупомянутого **метода замены ядра на вырожденное**, имеется ряд других приближенно-аналитических методов решения интегральных уравнений. Из них отметим методы, тесно примыкающие к аналогичным методам решения дифференциальных уравнений. Например, сродни рассмотренному в § 7.1 методу Пикара решения задачи Коши является **метод последовательных приближений** для интегральных уравнений; аналогом метода Галеркина решения краевой задачи (см. § 10.5) служит **метод моментов**. Описание этих и других приближенных методов можно найти в книгах [12, 19, 22, 31, 42, 55, 64].

Наиболее универсальными и хорошо приспособленными для компьютерных вычислений являются **численные** методы решения интегральных уравнений. Их построение опирается на замену интеграла в интегральном уравнении конечной суммой на базе какой-либо квадратурной формулы, в результате чего задача сводится к алгебраической системе относительно дискретных значений (каркаса) искомого решения, соответствующих заданным или определяемым выбором квадратурной формулы значениям аргумента (сетке). Такие методы называются **квадратурными методами** или **методами конечных сумм**. Они

чрезвычайно просты как в идее, так и в реализации, причем без каких-либо изменений их можно применить и к нелинейным интегральным уравнениям, имея лишь в виду, что в этом случае и системы конечномерных уравнений, к которым будет приводить дискретизация, также будут нелинейными.

Рассмотрению техники и особенностей применения метода конечных сумм к линейным интегральным уравнениям Фредгольма и Вольтерра посвящены следующие два параграфа.

§ 11.2. КВАДРАТУРНЫЙ МЕТОД РЕШЕНИЯ ИНТЕГРАЛЬНЫХ УРАВНЕНИЙ ФРЕДГОЛЬМА

Пусть для вычисления определенного интеграла используется некая конкретная квадратурная формула

$$\int_a^b \varphi(s) ds \approx \sum_{j=1}^n A_j \varphi(s_j) \quad (11.21)$$

с n узлами $s_j \in [a, b]$ и с соответствующими им весовыми коэффициентами A_j (см. гл.5).

Подставим правую часть приближенного равенства (11.21) с $\varphi(s) := Q(t, s)x(s)$ вместо интеграла в интегральное уравнение Фредгольма второго рода (11.5). В результате этого получаем

$$x(t) \approx \lambda \sum_{j=1}^n A_j Q(t, s_j) x(s_j) + f(t) \quad (11.22)$$

— приближенное представление решения $x(t)$ уравнения (11.5) через n его значений $x(s_1), x(s_2), \dots, x(s_n)$. Чтобы вычислить эти значения, станем рассматривать равенство (11.22) не при всех $t \in [a, b]$, а лишь на системе точек $t_1, t_2, \dots, t_n$, совпадающих соответственно с узлами $s_1, s_2, \dots, s_n$ квадратурной формулы (11.21). Таким образом, приходим к n равенствам вида

$$x(t_i) \approx \lambda \sum_{j=1}^n A_j Q(t_i, s_j) x(s_j) + f(t_i), \quad (11.23)$$

где $i = 1, 2, \dots, n$.

Положим для краткости

$$Q_{ij} := Q(t_i, s_j), \quad f_i := f(t_i), \quad x_i \approx x(t_i) (= x(s_i)). \quad (11.24)$$

Тогда (11.23) преобразуется к системе n линейных алгебраических уравнений с n неизвестными, которая выглядит следующим образом:

[illegible]

Стандартная запись линейной системы (11.25), с одной стороны, говорит о возможности ее решения тем или иным стандартным методом, если она совместна, с другой стороны, высвечивает роль параметра λ в однозначной разрешимости этой системы (и, как следствие, с некоторыми оговорками, интегрального уравнения (11.5)): при любых фиксированных A_j, Q_{ij} за счет уменьшения модуля параметра λ можно добиться диагонального преобладания в матрице системы, влекущего выполнимость условий сходимости метода простых итераций в форме Якоби [18].

Найдя решение системы (11.25), получаем n чисел $x_1, x_2, \dots, x_n$ — **каркас приближенного решения** интегрального уравнения на сетке $t = t_i$ ($i = 1, 2, \dots, n$). Если нам только и нужна эта таблица приближенных значений решения интегрального уравнения (11.5) в точках $t_1, t_2, \dots, t_n$, то в такой постановке задача решена. Если же нужны значения решения в других точках $t \in [a, b]$ или требуется знать решение $x(t)$ в аналитическом виде, подставляем найденные значения $x_j \approx x(s_j)$ ($j = 1, 2, \dots, n$) в равенство (11.22) или аппроксимируем их, привлекая для этих целей, например, интерполяцию, если n мало, или сплайн-интерполяцию при больших n ; в некоторых случаях здесь может оказаться полезным и метод наименьших квадратов.

Точность численного решения интегрального уравнения квадратурным методом зависит от нескольких взаимосвязанных факторов: применяемой квадратурной формулы, количества используемых узлов, свойств определяющих уравнение функций. Имеются неравенства, позволяющие, в принципе, оценить максимальное расхождение между точными и приближенными решениями при различных применяемых квадратурах (11.21) (см., например, [12, 19, 42, 55]); однако эти оценки малоприменимы в рядовых ситуациях ввиду громоздкости, что, в конечном итоге, отражается на их точности. Более практичным, хотя и недостаточно строгим способом контроля точности подсчитываемого методом квадратур каркаса решения, может служить уже использовавшийся ранее (см. §§ 5.5, 6.2, 7.7) принцип Рунге сравнения значений приближенных решений в общих узлах двух

разных сеток, используемых квадратурной формулой. Разумеется, при таком способе контроля точности приходится ограничиваться квадратурными формулами, опирающимися на равномерные сетки, или привлекать для сравнения результатов применения, например, формул Гаусса или Чебышева другие каркасы, подсчитывая значения приближенных решений в каких-то точках $t \in [a, b]$, отличных от узловых, по формуле (11.22).

При выборе конкретной квадратурной формулы (11.21) для использования в методе конечных сумм важно понимать, что лучшие результаты могут быть получены в том случае, когда ее алгебраический порядок точности хорошо согласуется со степенью гладкости ядра и свободного члена интегрального уравнения (предопределяющих гладкость искомого решения). Если ядро и свободный член не являются достаточно гладкими, то для дискретизации интегрального уравнения не следует применять высокоточных квадратур, а лучше ограничиться такими квадратурными формулами, как формулы трапеций или прямоугольников.

Последнее особенно существенно учитывать при решении интегральных уравнений первого рода (11.7), к которым формально можно применить тот же самый метод конечных сумм. Результаты при этом получаются гораздо менее надежными, в силу собственных свойств таких уравнений, что стимулировало развитие специального направления в вычислительной математике — методов решения некорректных задач или, иначе, методов регуляризации.

Пример 11.2. Дано уравнение

$$x(t) = \int_1^2 \frac{x(s) ds}{\sqrt{t+s^2}} + \sqrt{t+1} - \sqrt{t+4} + t. \quad (11.26)$$

К этому интегральному уравнению Фредгольма второго рода применим метод конечных сумм на основе квадратурной формулы Гаусса с двумя узлами (учитываем, что на отрезке $[1, 2]$ с гладкостью определяющих уравнение функций проблем нет). В соответствии с формулой (5.52) и табл.5.3 при $n = 2$ имеем

$$\int_1^2 \varphi(s) ds \approx \frac{1}{2} \left[\varphi\left(\frac{3-0.577350}{2}\right) + \varphi\left(\frac{3+0.577350}{2}\right) \right].$$

Используем это приближенное равенство в (11.26) с $\varphi(s) = \frac{x(s)}{\sqrt{t+s^2}}$. Приходим к приближенному представлению данного уравнения в виде

$$x(t) \approx \frac{x(1.211325)}{2\sqrt{t+1.211325^2}} + \frac{x(1.788675)}{2\sqrt{t+1.788675^2}} + \sqrt{t+1} - \sqrt{t+4} + t. \quad (11.27)$$

Полагая здесь аргумент t равным узловым значениям $t_1 = 1.211325$ и $t_2 = 1.788675$ (к которым мы пришли от узлов ∓ 0.577350 в результате сопоставления стандартного промежутка интегрирования $[-1, 1]$ с данным промежутком $[1, 2]$), получаем линейную систему

$$\begin{cases} 0.694498x_1 - 0.238077x_2 = 0.415545, \\ -0.277095x_1 + 0.776125x_2 = 1.052641 \end{cases}$$

типа системы (11.25) относительно неизвестных $x_1 \approx x(t_1)$, $x_2 \approx x(t_2)$. Решив эту систему, находим каркас приближенного решения

$$x_1 \approx 1.2116, \quad x_2 \approx 1.7888.$$

Восполнение этого каркаса по формуле (11.27) приводит к аналитическому виду приближенного решения уравнения (11.26)

$$x(t) \approx \tilde{x}(t) := \frac{0.6058}{\sqrt{t+1.4673}} + \frac{0.8944}{\sqrt{t+3.1994}} + \sqrt{t+1} - \sqrt{t+4} + t.$$

Чтобы получить представление о точности полученного приближенного решения, сравним его с известным точным решением $x(t) = t$ в узлах сетки (т.е. сравним каркасы, определяемые выбранной квадратурной формулой), а также в точках, служащих концами и серединой промежутка интегрирования (каркасы, представляющие потенциальный интерес). Результаты сравнения отражены следующей таблицей:

$t (= x(t))$	1	1.211325	1.5	1.788675	2
$\tilde{x}(t)$	1.0003	1.2116	1.5002	1.7888	2.0001

В случае применения простейшей формулы Симпсона (5.19), опирающейся здесь на узлы $t_1 = 1$, $t_2 = 1.5$, $t_3 = 2$, приходим к представлению решения

$$x(t) \approx \frac{1}{6} \left[\frac{x(1)}{\sqrt{t+1}} + \frac{4x(1.5)}{\sqrt{t+2.25}} + \frac{x(2)}{\sqrt{t+4}} \right] + \sqrt{t+1} - \sqrt{t+4} + t$$

и соответственно к линейной алгебраической системе

$$\begin{cases} 0.88215x_1 - 0.36980x_2 - 0.07454x_3 &= 0.17815, \\ -0.10541x_1 + 0.65574x_2 - 0.07107x_3 &= 0.73593, \\ -0.09623x_1 - 0.32338x_2 + 0.93196x_3 &= 1.28256. \end{cases}$$

Ее решение дает значения

$$x_1 \approx 0.9996, \quad x_2 \approx 1.4997, \quad x_3 \approx 1.9998,$$

также являющиеся достаточно хорошими приближениями к точным значениям $x(1) = 1$, $x(1.5) = 1.5$, $x(2) = 2$.

Замечание 11.1. При численном решении интегральных уравнений с разрывными ядрами или с бесконечными границами интегрирования следует вспомнить о наличии специально приспособленных для таких случаев квадратурных формул Гаусса–Кристоффеля, описанию которых посвящен § 5.7. Сделав правильный выбор квадратурной формулы из указанного семейства, все остальное можно выполнять по рассмотренной выше схеме.

§ 11.3. КВАДРАТУРНЫЙ МЕТОД РЕШЕНИЯ ИНТЕГРАЛЬНЫХ УРАВНЕНИЙ ВОЛЬТЕРРА

Выше отмечалось, что параметр λ в линейных интегральных уравнениях Вольтерра не несет такой нагрузки, как в случае уравнений Фредгольма. Поэтому положим в уравнении (11.6) $\lambda = 1$ и будем численно решать уравнение

$$x(t) = \int_a^t Q(t, s)x(s)ds + f(t), \quad t \in [a, b]. \quad (11.28)$$

Помня о том, что это уравнение формально можно считать уравнением Фредгольма вида

$$x(t) = \int_a^b K(t, s)x(s)ds + f(t) \quad (11.29)$$

с ядром

$$K(t, s) = \begin{cases} Q(t, s) & \text{при } a \leq s \leq t \leq b, \\ 0 & \text{при } a \leq t \leq s \leq b, \end{cases} \quad (11.30)$$

для приближенного представления его решения

$$x(t) \approx \sum_{j=1}^n A_j Q(t, s_j)x(s_j) + f(t) \quad (11.31)$$

на основе квадратурной формулы с узлами $s_j \in [a, b]$ и весами A_j можно воспользоваться результатами предыдущего параграфа. Согласно им, для получения каркаса $x_1, x_2, \dots, x_n$ приближенного решения (11.31) нужно составить систему линейных алгебраических уравнений типа системы (11.25), которая применительно к уравнению (11.29), в силу (11.30), превращается

в треугольную:

[illegible]

Отсюда легко получаем последовательно один за другим иско-
мые значения $x_1, x_2, \dots, x_n$, полагая $i = 1, 2, \dots, n$ в формуле

$$x_i = \frac{f_i + \sum_{j=1}^{i-1} A_j Q_{ij} x_j}{1 - A_i Q_{ii}}. \quad (11.33)$$

Применение такого формального подхода к численному решению уравнений Вольтерра имеет некоторые нюансы. Чтобы не разбираться с ними на этой стадии решения рассматриваемой задачи, проще заняться процессом дискретизации интегрального уравнения (11.28) с самого начала. При этом, учитывая, что точка b , ограничивающая промежуток определения решения $x(t)$, в уравнениях Вольтерра присутствует лишь номинально (как мы помним, эти уравнения родственны начальным задачам для ОДУ), то при их дискретизации лучше изменить очередность фиксирования переменных t и s по сравнению со случаем уравнений Фредгольма.

Будем придавать аргументу t в уравнении (11.28) возрастающие значения $t_1, t_2, \dots, t_n \in [a, b]$. Получим n интегральных равенств

$$x(t_i) = \int_a^{t_i} Q(t_i, s)x(s)ds + f(t_i), \quad (11.34)$$

промежутки интегрирования в которых при каждом i теперь постоянны и увеличиваются с ростом i на величину шага $h_i := t_i - t_{i-1}$. Заменяв определенные интегралы в (11.34) конечными суммами

$$\sum_{j=1}^i A_j Q(t_i, s_j) x(s_j) \quad (11.35)$$

с переменным числом узлов, равным i^*), приходим к

^{*)} Что можно отразить введением второго индекса в обозначение весового коэффициента квадратурной формулы, т.е. всюду в пределах этого параграфа есть смысл использовать A_{ij} в роли A_j

приближенным равенствам

$$x(t_i) \approx \sum_{j=1}^i A_j Q(t_i, s_j) x(s_j) + f(t_i). \quad (11.36)$$

Если, как и прежде, при любых $i \in \{1, 2, \dots, n\}$

$$t_i = s_i \quad \text{и} \quad x_i \approx x(t_i),$$

то (11.36) определяет линейную алгебраическую систему

$$x_i = \sum_{j=1}^i A_j Q_{ij} x_j + f_i, \quad i = 1, 2, \dots, n, \quad (11.37)$$

в точности совпадающую с формально выписанной выше системой (11.32).

Из последних рассуждений, в частности, вытекает понимание того, что совокупность всех весовых коэффициентов A_j в каждой строке системы (11.32) должна быть полной для выбранной квадратурной формулы, т.е. при каждом $i = 1, 2, \dots, n$ должно иметь место равенство

$$\sum_{j=1}^i A_j = t_i - a.$$

При конкретизации вида квадратурной формулы для замены определенных интегралов в равенствах (11.34) конечными суммами (11.35) предпочтение здесь отдается формулам замкнутого типа, например, квадратурной формуле трапеций (не обязательно с постоянным шагом h_i). В таком случае при $t_1 = a$ из (11.34) сразу следует

$$x_1 = f_1 \quad (= f(a)).$$

Далее, при $i = 2$ к интегралу $\int_a^{t_2} = \int_a^{t_2}$ в (11.34) применяется простейшая формула трапеций (5.17), в результате чего получаем второе уравнение системы (11.37):

$$x_2 = \frac{t_2 - a}{2} (Q_{21} x_1 + Q_{22} x_2) + f_2.$$

К следующему интегралу $\int_a^{t_3} = \int_a^{t_3}$ применяется уже составная формула трапеций (5.28), которая в более простом варианте с

постоянным шагом $h = t_i - t_{i-1}$ приводит к уравнению

$$x_3 = \frac{h}{2}Q_{31}x_1 + hQ_{32}x_2 + \frac{h}{2}Q_{33}x_3 + f_3,$$

и так далее. Таким образом, в случае использования квадратурной формулы трапеций с постоянным шагом $h = t_i - t_{i-1}$ ($i = 2, 3, \dots, n$) треугольная система для получения каркаса $x_1, x_2, \dots, x_n$ приближенного решения линейного интегрального уравнения Вольтерра второго рода (11.28) приобретает вид^{*})

$$\left\{ \begin{array}{l} x_1 = f_1, \\ -\frac{h}{2}Q_{21}x_1 + \left(1 - \frac{h}{2}Q_{22}\right)x_2 = f_2, \\ -\frac{h}{2}Q_{31}x_1 - hQ_{32}x_2 + \left(1 - \frac{h}{2}Q_{33}\right)x_3 = f_3, \\ \dots\dots\dots \\ -\frac{h}{2}Q_{n1}x_1 - hQ_{n2}x_2 - \dots - hQ_{n,n-1}x_{n-1} + \\ \qquad \qquad \qquad + \left(1 - \frac{h}{2}Q_{nn}\right)x_n = f_n. \end{array} \right. \quad (11.38)$$

Нет никакой необходимости собирать сразу все требуемые для построения каркаса решения уравнения в систему, целесообразнее поочередно записывать их и по ходу разрешать относительно соответствующего неизвестного или просто последовательно пользоваться формулой для вычислений неизвестных $x_2, x_3, \dots$ типа формулы (11.33).

Пример 11.3. Дано интегральное уравнение Вольтерра второго рода

$$x(t) = \int_0^t t \cos^2(ts^3)x(s)ds + t^2 - \frac{1}{3}tg(t^4). \quad (11.39)$$

Применим к нему квадратурную формулу трапеций с шагом $h = 0.1$, используя четыре узла:

$$t_1 = s_1 = 0, \quad t_2 = s_2 = 0.1, \quad t_3 = s_3 = 0.2, \quad t_4 = s_4 = 0.3.$$

В соответствии с этой сеткой, рис.11.2, и видом отвечающих данному

^{*)} В случае переменного шага ее вид незначительно сложнее.

случаю уравнений (11.38) имеем:

$$\begin{aligned}
 x_1 &= f(0) = 0, \\
 x_2 &= \frac{0.01 - \frac{1}{3} \operatorname{tg} 0.0001}{1 - 0.005 \cdot \cos^2 0.0001} \approx 0.010017, \\
 x_3 &\approx \frac{0.04 - \frac{1}{3} \operatorname{tg} 0.0016 + 0.02 \cos^2 0.0002 \cdot 0.010017}{1 - 0.01 \cos^2 0.0016} \approx 0.040068, \\
 x_4 &\approx \frac{0.09 - \frac{1}{3} \operatorname{tg} 0.0081 + 0.03 \cos^2 0.0003 \cdot 0.010017}{1 - 0.015 \cos^2 0.0081} + \\
 &\quad + \frac{0.03 \cos^2 0.0024 \cdot 0.040068}{1 - 0.015 \cos^2 0.0081} \approx 0.090155.
 \end{aligned}$$

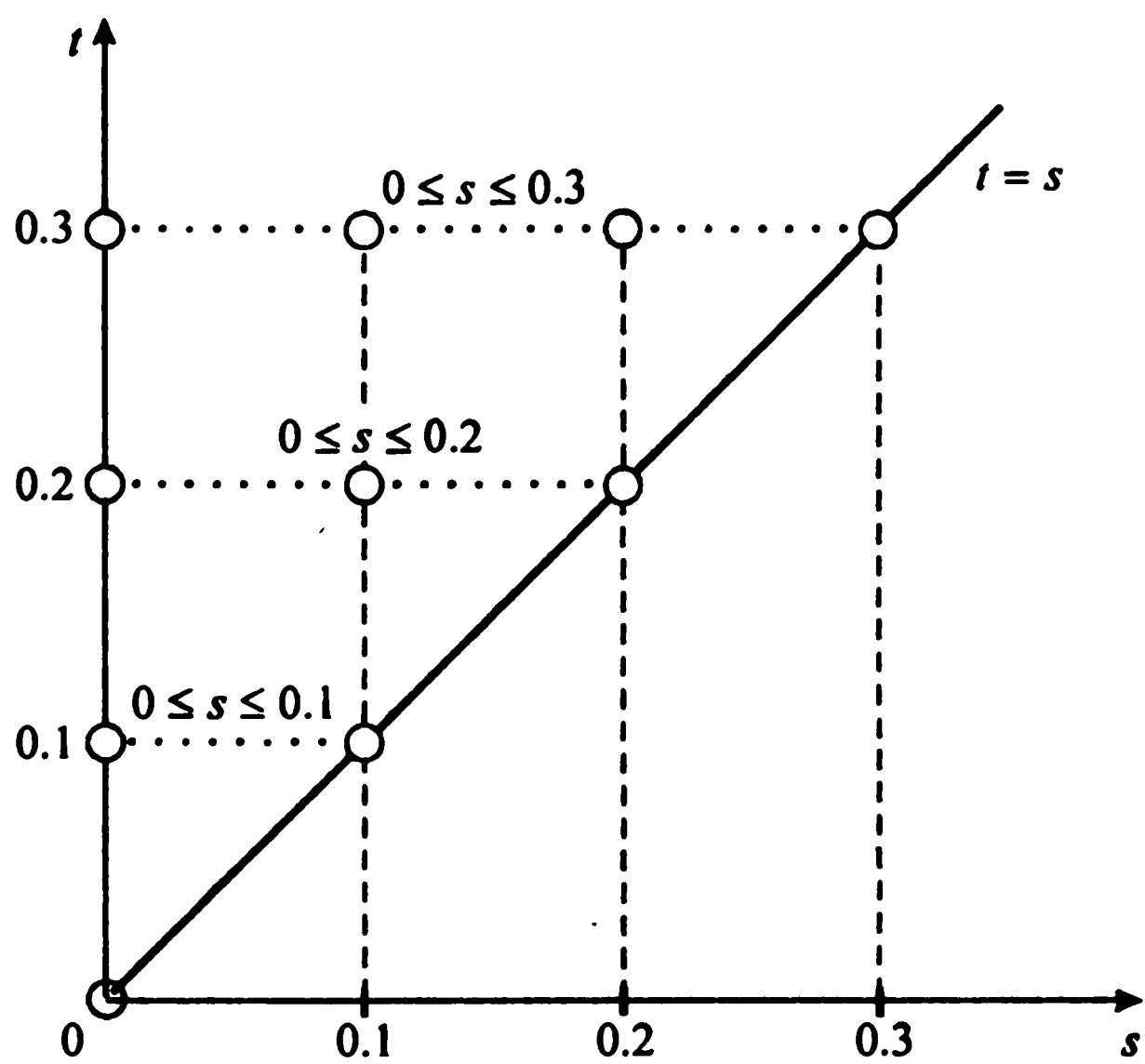


Рис. 11.2. Узлы и промежутки последовательного интегрирования при численном решении уравнения (11.39) методом трапеций с шагом $h = 0.1$

Полученные значения x_1, x_2, x_3, x_4 хорошо согласуются с соответствующими значениями $x(0), x(0.1), x(0.2), x(0.3)$ точного решения данного уравнения (11.39) $x(t) = t^2$. При этом замечаем, что с удалением от точки $a = 0$ точек t_i — фиксированных верхних границ интегрирования в выражениях вида (11.34) — точность уменьшается, что характерно и при численном нахождении решений задач Коши для ОДУ с постоянным шагом при удалении от начальной точки. Так как перемена очередности фиксирования переменных t и s в интегральном уравнении Вольтерра лишила нас аналитического представления приближенного решения через его каркас типа представления (11.22), применим лагранжеву интерполяцию. Составим по найденным числам x_1, x_2, x_3, x_4 — приближенным значениям решения $x(t)$ — таблицу конечных разностей:

i	t_i	x_i	Δx_i	$\Delta^2 x_i$	$\Delta^3 x_i$
1	0	0	0.010017		
2	0.1	0.010017	0.030051	0.020034	0.000002
3	0.2	0.040068	0.050087	0.020036	
4	0.3	0.090155			

Вторые разности в ней практически совпадают, поэтому следует ограничиться квадратичной интерполяцией. Взяв за основу первые три узла, по первой формуле Ньютона (1.25) получаем

$$\begin{aligned} x(t) \approx P_2(t) &= x_1 + \frac{\Delta x_1}{h}(t - t_1) + \frac{\Delta^2 x_1}{2h^2}(t - t_1)(t - t_2) = \\ &= 0 + \frac{0.010017}{0.1}t + \frac{0.020034}{2 \cdot 0.01}t(t - 0.1) = 1.0017t^2. \end{aligned}$$

По трем следующим узлам (t_2, t_3, t_4) аналогично находим

$$x(t) \approx \tilde{P}_2(t) = 1.0018t^2 - 0.00003t + 0.000002.$$

И то, и другое выражения могут служить неплохими аппроксимациями точного решения $x(t)$. Можно рассчитывать на лучшие результаты, если рассматривать сетку с более мелким шагом h .

Для решения уравнений Вольтерра первого рода (11.8) удобнее применять квадратурные формулы открытого типа, например, формулу средних прямоугольников. Фиксируя в таком уравнении переменную t равной значениям

$$t_1 = a + h, \quad t_2 = t_1 + h, \quad \dots, \quad t_i = t_{i-1} + h, \tag{11.40}$$

получаем равенства

$$\int_a^{t_i} Q(t_i, s)x(s)ds = f(t_i), \quad i = 1, 2, \dots, \quad (11.41)$$

из которых видно, что здесь нет необходимости считать узлы квадратур s_i совпадающими с узлами t_i ; для формулы прямоугольников (средней точки) (5.8) берем их посередине элементарных промежутков интегрирования $[t_{i-1}, t_i]$ (рис 11.3).

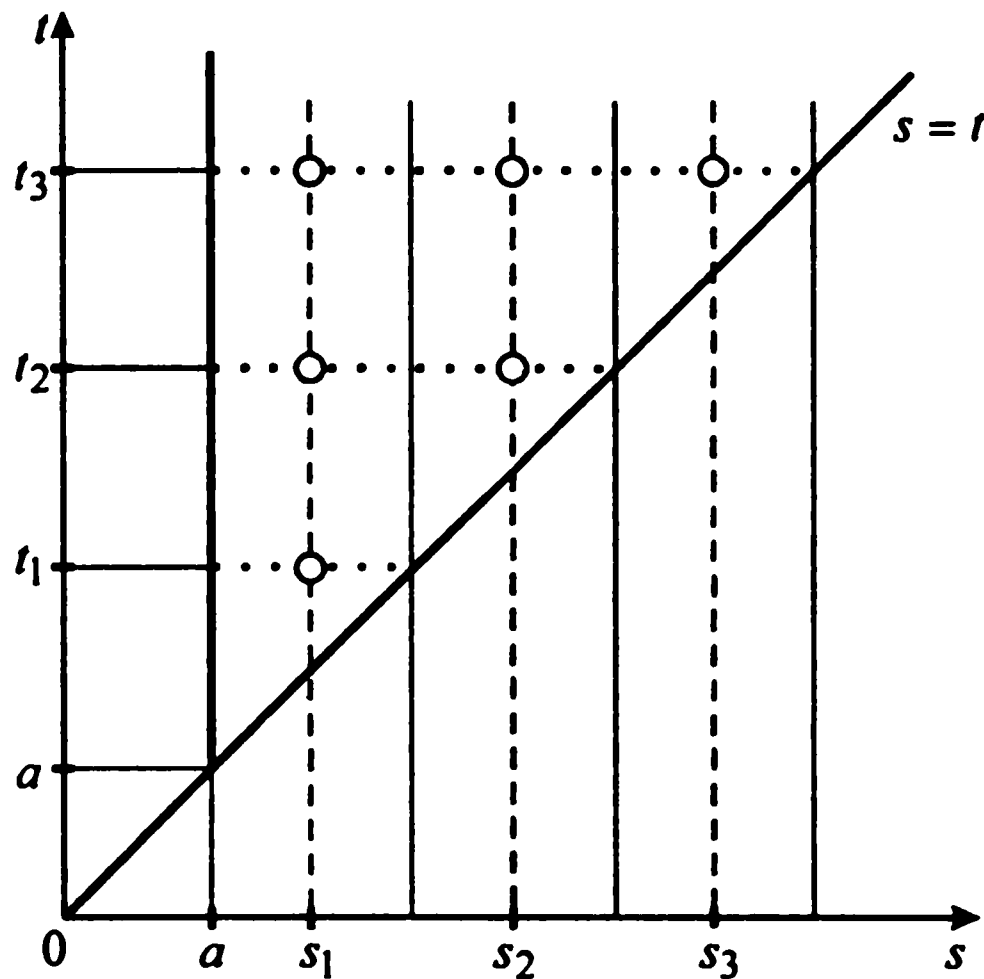


Рис. 11.3. Расположение узлов квадратурных формул прямоугольников при численном решении уравнений Вольтерра

При этом каркас приближенного решения определяется именно узлами s_i , т.е. полагаем

$$x_1 \approx x(s_1), \quad x_2 \approx x(s_2), \quad \dots, \quad x_n \approx x(s_n). \quad (11.42)$$

При взятой за основу квадратурной формуле прямоугольников с учетом (11.42) равенства (11.41) приводят к следующим:

$$\begin{aligned} hQ(t_1, s_1)x_1 &= f(t_1), \\ hQ(t_2, s_1)x_1 + hQ(t_2, s_2)x_2 &= f(t_2), \\ hQ(t_3, s_1)x_1 + hQ(t_3, s_2)x_2 + hQ(t_3, s_3)x_3 &= f(t_3), \end{aligned} \quad (11.43)$$

и т.д. Напомним, что здесь узлы $t_1, t_2, \dots$ сетки на оси Ot определены в (11.40), а

$$s_i := t_i - \frac{h}{2} \quad \forall \quad i = 1, 2, \dots$$

Из равенств (11.43) последовательно находим:

$$\begin{aligned} x_1 &= \frac{f(t_1)}{hQ(t_1, s_1)}, \\ x_i &= \frac{f(t_i) - h \sum_{j=1}^{i-1} Q(t_i, s_j)x_j}{hQ(t_i, s_i)} \quad (i = 2, 3, \dots). \end{aligned} \quad (11.44)$$

В случае применения в (11.41) квадратурных формул замкнутого типа при совпадающих системах узлов $\{s_i\}$ и $\{t_i\}$ возникает проблема вычисления значения $x_1 \approx x(t_1) = x(s_1) = x(a)$, которой не было в аналогичной ситуации с уравнениями второго рода. Действительно, это значение не может быть найдено непосредственно из равенства (11.41), при $i = 1$ теряющего смысл, а без него нельзя подсчитывать последующие значения $x_2, x_3, \dots$

Чтобы вычислить x_1 , продифференцируем рассматриваемое уравнение (11.8) по t и в полученном таким образом уравнении второго рода

$$Q(t, t)x(t) + \int_a^t Q'_t(t, s)x(s)ds = f'(t)$$

положим $t = a$. Имеем равенство:

$$Q(a, a)x(a) = f'(a),$$

откуда следует, что $x(a) = f'(a)/Q(a, a)$, т.е. можно принять

$$x_1 = \frac{f'(a)}{Q_{11}}. \quad (11.45)$$

Далее при использовании, например, квадратурной формулы

трапеций следует:

$$\frac{h}{2}Q_{21}x_1 + \frac{h}{2}Q_{22}x_2 = f_2 \Rightarrow x_2 = \frac{f_2 - \frac{h}{2}Q_{21}x_1}{\frac{h}{2}Q_{22}}, \quad (11.46)$$

$$\begin{aligned} \frac{h}{2}Q_{31}x_1 + hQ_{32}x_2 + \frac{h}{2}Q_{33}x_3 = f_3 \Rightarrow \\ x_3 = \frac{f_3 - \frac{h}{2}Q_{31}x_1 - hQ_{32}x_2}{\frac{h}{2}Q_{33}}, \end{aligned} \quad (11.47)$$

т.е. в общем случае при любом $j = 2, 3, \dots$

$$x(s_j) \approx x_j = \frac{f_j - \frac{h}{2}Q_{j1}x_1 - h \sum_{k=2}^{j-1} Q_{jk}x_k}{\frac{h}{2}Q_{jj}} \quad (11.48)$$

(обозначения те же, что и в (11.24)).

Пример 11.4. Рассмотрим численное решение уравнения

$$\int_1^t (t^2 + s^2 + 1) x(s) ds = t^2 - \frac{1}{t} \quad (11.49)$$

на промежутке $[1, 1.3]$ по формулам трапеций и прямоугольников, полагая шаг $h = 0.1$. (Точное решение $x(t) = \frac{1}{t^2}$ со значениями

$$x(1) = 1, \quad x(1.1) \approx 0.82645, \quad x(1.2) \approx 0.6944, \quad x(1.3) \approx 0.59172,$$

$$x(1.05) \approx 0.90703, \quad x(1.15) \approx 0.75614, \quad x(1.25) \approx 0.64).$$

Применим сначала формулу трапеций. Зафиксируем значения

$$t_1 = 1, \quad t_2 = 1.1, \quad t_3 = 1.2, \quad t_4 = 1.3$$

и в получающихся при этом из (11.49) равенствах

$$\int_1^{t_i} (t_i^2 + s^2 + 1) x(s) ds = t_i^2 - \frac{1}{t_i}, \quad i = 1, 2, 3, 4$$

узлами s_j квадратур считаем те же точки 1, 1.1, 1.2, 1.3. Вычислив по формуле (11.45) начальное значение решения

$$x(1) = x_1 = \frac{f'(1)}{Q(1, 1)} = \frac{2t + \frac{1}{t^2}}{2t^2 + 1} \bigg|_{t=1} = 1,$$

последующие его значения на сетке $\{s_j\}$ находим по формулам (11.46)–(11.48). Имеем:

$$x(1.1) \approx x_2 = \frac{1.1^2 - \frac{1}{1.1} - 0.05(1.1^2 + 1^2 + 1) \cdot 1}{0.05(1.1^2 + 1.1^2 + 1)} \approx 0.8211,$$

$$x(1.2) \approx x_3 = \frac{1.2^2 - \frac{1}{1.2} - 0.05(1.2^2 + 1^2 + 1) \cdot 1 - 0.1 \cdot (1.2^2 + 1.1^2 + 1) \cdot 0.8211}{0.05(1.2^2 + 1.2^2 + 1)} \approx 0.6957,$$

$$x(1.3) \approx x_4 = \frac{1.3^2 - \frac{1}{1.3} - 0.05(1.3^2 + 1^2 + 1) \cdot 1 - 0.1(1.3^2 + 1.1^2 + 1) \cdot 0.8211}{0.05(1.3^2 + 1.3^2 + 1)} - \frac{0.1(1.3^2 + 1.2^2 + 1) \cdot 0.6957}{0.05(1.3^2 + 1.3^2 + 1)} \approx 0.5877.$$

Теперь обратимся к формуле прямоугольников. Для данного случая в равенства (11.43) или в вытекающие из них формулы (11.44) при заданном $h = 0.1$ нужно подставлять следующие значения:

$$Q(t_1, s_1) = Q(1.1, 1.05) = 3.3125, \quad f(t_1) = f(1.1) \approx 0.30091;$$

$$Q(t_2, s_1) = Q(1.2, 1.05) = 3.5425,$$

$$Q(t_2, s_2) = Q(1.2, 1.15) = 3.7625, \quad f(t_2) = f(1.2) \approx 0.60667;$$

$$Q(t_3, s_1) = Q(1.3, 1.05) = 3.7925,$$

$$Q(t_3, s_2) = Q(1.3, 1.15) = 4.0125,$$

$$Q(t_3, s_3) = Q(1.3, 1.25) = 4.2525, \quad f(t_3) = f(1.3) \approx 0.92077.$$

Пользуясь ими, последовательно вычисляем:

$$x(1.05) \approx x_1 = \frac{0.30091}{0.1 \cdot 3.3125} \approx 0.90841,$$

$$x(1.15) \approx x_2 = \frac{0.60667 - 0.1 \cdot 3.5425 \cdot 0.90841}{0.1 \cdot 3.7625} \approx 0.75712,$$

$$x(1.25) \approx x_3 = \frac{0.92077 - 0.1 \cdot 3.7925 \cdot 0.90841 - 0.1 \cdot 4.0125 \cdot 0.75712}{0.1 \cdot 4.2525} \approx 0.64071.$$

Сравнение тех и других серий приближенных значений решения с точными значениями демонстрирует достаточно высокую эффективность применяемых квадратур (как и следовало ожидать, несколько лучшие результаты показывает квадратурная формула средней точки).

§ 11.4. КВАДРАТУРНО-ИТЕРАЦИОННЫЙ МЕТОД ПОСТРОЕНИЯ РЕЗОЛЬВЕНТ

Предположим, что мы находимся в ситуации, когда требуется решать серию интегральных уравнений второго рода с одним и тем же ядром $Q(t, s)$ и разными свободными членами $f(t)$ ^{*)}. Для определенности будем считать, что речь идет об интегральном уравнении Фредгольма (11.5). В таком случае, подобно тому, как это делается для линейных алгебраических систем $Ax = b$, где решения x^* при разных правых частях b вычисляются через предварительно найденную обратную к A матрицу A^{-1} по формуле $x^* = A^{-1}b$, здесь также имеется возможность формально подсчитывать соответствующие разным $f(t)$ решения $x^*(t)$ уравнения (11.5) по формуле вида

$$x^*(t) = \lambda \int_a^b R(t, \tau; \lambda) f(\tau) d\tau + f(t). \quad (11.50)$$

В этой формуле роль разрешающего оператора (аналога обратному) играет функция $R(t, s; \lambda)$, называемая *резольвентой* интегрального уравнения или, быть может, точнее, резольвентой семейства ядер $\lambda Q(t, s)$ уравнения (11.5)^{**)} . При непрерывных

^{*)} Оговоренная ситуация возникает, например, при решении нелинейного интегрального уравнения модифицированным методом Ньютона [41, 51].

^{**)} Иногда вместо термина *резольвента* используют термин *разрешающее ядро* [59].

ядрах $Q(t, s)$ и значениях параметра λ , удовлетворяющих, например, условию (11.10), несложно установить существование единственной непрерывной функции $R(t, s; \lambda)$, к которой абсолютно и равномерно при $t, s \in [a, b]$ сходится *ряд Неймана*^{*)}

$$\sum_{k=1}^{\infty} \lambda^{k-1} Q_k(t, s), \quad (11.51)$$

определяемый через *итерированные ядра* $Q_k(t, s)$ такие, что

$$Q_1(t, s) \equiv Q(t, s), \quad Q_k(t, s) = \int_a^b Q(t, \tau) Q_{k-1}(\tau, s) d\tau. \quad (11.52)$$

Последовательное вычисление итерированных ядер по рекуррентной формуле (11.52) и подсчет частичных сумм ряда (11.51) позволяют, в принципе, получить сколь угодно хорошее приближение к резольвенте $R(t, s; \lambda)$. Но цена такого приближения будет слишком высока (за редкими исключениями, когда такую процедуру можно провести аналитически, а не численно). Более продуктивные способы приближенного нахождения резольвенты опираются на простые интегральные соотношения между резольвентой и ядром. Практически в любом пособии по интегральным уравнениям можно найти вывод таких соотношений; их вид

$$R(t, s; \lambda) = Q(t, s) + \lambda \int_a^b Q(t, \tau) R(\tau, s; \lambda) d\tau \quad (11.53)$$

и

$$R(t, s; \lambda) = Q(t, s) + \lambda \int_a^b R(t, \tau, \lambda) Q(\tau, s) d\tau. \quad (11.54)$$

Тот факт, что резольвента, согласно формулам (11.53), (11.54), может рассматриваться как решение интегрального уравнения с тем же ядром $Q(t, s)$ (при фиксировании в ней одного из двух аргументов), позволяет подходить к ее численному построению с тех же позиций, которые были заложены в параграфе 11.2 при приближенном решении уравнений Фредгольма второго рода (11.5).

Остановившись на какой-либо квадратурной формуле вида (11.21), по ее узлам на отрезке $[a, b]$ и, соответственно, на квадрате $[a, b] \times [a, b]$ строим сетку, узлами которой служат точки $(t_i; s_j)$, лежащие на пересечении линий $t = t_i, s = s_j$

^{*)} Нейман Карл Готфрид (1832–1925) — немецкий математик. Его основные труды относятся к дифференциальным уравнениям и алгебраическим функциям.

$(i, j = 1, 2, \dots, n)$ в пределах данного квадрата^{*)}. Применение выбранной квадратурной формулы к интегральному соотношению (11.54) приводит к равенству

$$R(t, s; \lambda) = Q(t, s) + \lambda \sum_{k=1}^n A_k R(t, \tau_k; \lambda) Q(\tau_k, s) + r(t, s),$$

где $r(t, s)$ — погрешность квадратуры. Отбросив эту погрешность (малости которой можно добиться разными способами, например, увеличением числа узлов n), переходим к приближенному уравнению относительно точной резольвенты $R(t, s; \lambda)$, которое, в свою очередь, заменяем точным уравнением относительно приближенной резольвенты $\tilde{R}(t, s; \lambda) \approx R(t, s; \lambda)$:

$$\tilde{R}(t, s; \lambda) = Q(t, s) + \lambda \sum_{k=1}^n A_k \tilde{R}(t, \tau_k; \lambda) Q(\tau_k, s). \quad (11.55)$$

Из функционального уравнения (11.55) фиксированием переменных $t = t_i, s = s_j$ ($i, j = 1, 2, \dots, n$) получаем n^2 дискретных уравнений

$$\tilde{R}(t_i, s_j; \lambda) = Q(t_i, s_j) + \lambda \sum_{k=1}^n A_k \tilde{R}(t_i, \tau_k; \lambda) Q(\tau_k, s_j) \quad (11.56)$$

относительно сеточных значений приближенной резольвенты, т.е. ее *каркаса* $\tilde{R}_{ij}(\lambda) := \tilde{R}(t_i, s_j; \lambda)$.

Введем $n \times n$ -матрицы

$$\hat{\mathbf{R}}(\lambda) = (\tilde{R}(t_i, s_j; \lambda))_{i,j=1}^n, \quad \hat{\mathbf{Q}} := (Q(t_i, s_j))_{i,j=1}^n, \\ \mathbf{A} := \text{diag}(A_k)_{k=1}^n.$$

Учитывая совпадение значений t_i, s_j, τ_k при совпадении индексов i, j, k , совокупность уравнений (11.56) теперь можно переписать в виде одного матричного уравнения

$$\hat{\mathbf{R}}(\lambda) = \hat{\mathbf{Q}} + \lambda \hat{\mathbf{R}}(\lambda) \mathbf{A} \hat{\mathbf{Q}}, \quad (11.57)$$

в чем нетрудно убедиться, выполняя непосредственно (в элементах) фигурирующие в (11.57) матричные операции.

Формально матричное уравнение (11.57) легко разрешается

^{*)} Можно пойти и обратным путем: сначала ввести сетку, например, равномерную, а уже к ней подобрать подходящую квадратурную формулу.

относительно искомого каркаса приближенной резольвенты:

$$\hat{R}(\lambda) = \hat{Q}(\mathbf{E} - \lambda \mathbf{A} \hat{Q})^{-1}. \quad (11.58)$$

Согласно лемме Неймана, матрица $(\mathbf{E} - \lambda \mathbf{A} \hat{Q})^{-1}$ существует и представима матричным рядом

$$(\mathbf{E} - \lambda \mathbf{A} \hat{Q})^{-1} = \mathbf{E} + \lambda \mathbf{A} \hat{Q} + \lambda^2 (\mathbf{A} \hat{Q})^2 + \lambda^3 (\mathbf{A} \hat{Q})^3 + \dots \quad (11.59)$$

в том и только в том случае, когда спектральный радиус матрицы $\lambda \mathbf{A} \hat{Q}$ меньше единицы. Предположим, что это условие выполнено, т.е.

$$\rho(\lambda \mathbf{A} \hat{Q}) < 1. \quad (11.60)$$

Тогда искомый каркас приближенной резольвенты, в силу (11.58) и (11.59), также можно записать в виде суммы матричного ряда:

$$\hat{R}(\lambda) = \hat{Q} + \lambda \hat{Q} \mathbf{A} \hat{Q} + \lambda^2 \hat{Q} (\mathbf{A} \hat{Q})^2 + \lambda^3 \hat{Q} (\mathbf{A} \hat{Q})^3 + \dots \quad (11.61)$$

Вместо получения приближений к $\hat{R}(\lambda)$ подсчетом частичных сумм ряда (11.61), который может сходиться весьма медленно, вернемся к представлению (11.58) и для обращения матрицы $\mathbf{E} - \lambda \mathbf{A} \hat{Q}$ применим быстросходящийся итерационный процесс Шульца [18].

Чтобы не усложнять записи, примем за основу наиболее простой и распространенный процесс уточнения элементов обратной матрицы $\mathbf{B}^{-1}$ по формулам

$$\begin{cases} \mathbf{U}_{k+1} = \mathbf{U}_k + \mathbf{U}_k \boldsymbol{\Psi}_k, \\ \boldsymbol{\Psi}_k = \mathbf{E} - \mathbf{B} \mathbf{U}_k, \quad k = 0, 1, 2, \dots \end{cases} \quad (11.62)$$

Известно [18], что необходимым и достаточным условием сходимости последовательности матриц $\mathbf{U}_k$ к матрице $\mathbf{B}^{-1}$ является условие

$$\rho(\boldsymbol{\Psi}_0) < 1, \quad (11.63)$$

и если начальное приближение $\mathbf{U}_0$ выбрано так, что

$$\|\boldsymbol{\Psi}_0\| \leq q < 1,$$

то справедливы оценки погрешности

$$\|\mathbf{B}^{-1} - \mathbf{U}_k\| \leq \frac{\|\mathbf{U}_k \boldsymbol{\Psi}_k\|}{1 - \|\boldsymbol{\Psi}_k\|} \leq \frac{\|\mathbf{U}_0\|}{1 - q} q^{2^k}, \quad (11.64)$$

характеризующие (11.62) как метод второго порядка.

При обращении матрицы $E - \lambda A \hat{Q}$ в записи итерационного процесса (11.62) изменяется лишь вторая строка (для подсчета невязки), т.е. приближения U_k к $(E - \lambda A \hat{Q})^{-1}$ можно находить по формулам

$$\begin{cases} U_{k+1} = U_k + U_k \psi_k, & k = 0, 1, 2, \dots, \\ \psi_k = E - (E - \lambda A \hat{Q}) U_k, \end{cases} \quad (11.65)$$

если выполняется условие (11.63). Учитывая, что при условии (11.60) матрица $(E - \lambda A \hat{Q})^{-1}$ может быть разложена в ряд по формуле (11.59), возьмем в качестве начальной матрицы U_0 в процессе (11.65) первый член этого разложения, т.е. положим

$$U_0 := E.$$

Тогда начальная невязка, «величина» которой определяет сходимость метода, будет

$$\psi_0 = E - (E - \lambda A \hat{Q}) E = \lambda A \hat{Q},$$

и, следовательно, в таком случае необходимое и достаточное условие (11.63) сходимости итерационного процесса (11.65) совпадает с необходимым и достаточным условием (11.60) существования искомой обратной матрицы и ее представимости матричным рядом (11.59).

Таким образом, если $\rho(\lambda A \hat{Q}) < 1$, то начатый с $U_0 = E$ итерационный процесс (11.65) сходится к матрице $(E - \lambda A \hat{Q})^{-1}$, а значит, и последовательность матриц $R_k := \hat{Q} U_k$, в силу равенства (11.58), сходится к матрице $\hat{R}(\lambda)$ — каркасу приближенной резольвенты $\tilde{R}(t, s; \lambda)$ при всяком фиксированном λ , отвечающем условию $\rho(\lambda A \hat{Q}) < 1$. Если, более того,

$$\|\lambda A \hat{Q}\| \leq q < 1, \quad (11.66)$$

то можно воспользоваться вытекающими из (11.64) оценками (апостериорной и априорной)

$$\|\hat{R}(\lambda) - R_k\| \leq \|\hat{Q}\| \frac{\|U_k \psi_k\|}{1 - \|\psi_k\|} \leq \frac{\|\hat{Q}\|}{1 - q} q^{2^k} \quad (11.67)$$

при любых мультипликативных нормах таких, что $\|E\| = 1$.

Подводя итог, скажем, что при определенных условиях, например, за счет измельчения сетки, искомая резольвента

$R(t, s; \lambda)$ уравнения Фредгольма второго рода с ядром $\lambda Q(t, s)$ при фиксированных значениях λ может быть сколь угодно хорошо представлена функцией $\tilde{R}(t, s; \lambda)$, т.е. $n \times n$ -матрицей $\hat{R}(\lambda)$; последняя, в свою очередь, может быть сколь угодно хорошо приближена матрицами $R_k := \hat{Q}U_k$ с помощью квадратично сходящегося итерационного процесса (11.65) вычисления матриц U_k (предварительно проверяется условие (11.66)).

Пример 11.5. Рассмотрим поведение описанного процесса построения приближений к каркасу резольвенты, взяв ядро

$$Q(t, s) = t^2 s - t s^2, \quad 0 \leq t, s \leq 1 \quad (11.68)$$

и параметр $\lambda = 4$.

Воспользуемся квадратурной формулой Гаусса с двумя узлами

$$\int_0^1 \varphi(x) dx \approx 0.5\varphi(0.211325) + 0.5\varphi(0.788675)$$

(см. формулу (5.52) при $a = 0$, $b = 1$, $n = 2$) и в соответствии с ней зададим на квадрате $[0, 1] \times [0, 1]$ 2×2 -сетку узлов $(t_i; s_j)$ искомого каркаса так, как это показано на рис. 11.4, где для удобства сопоставления сеточных значений и элементов матриц ось Ot «перевернута».

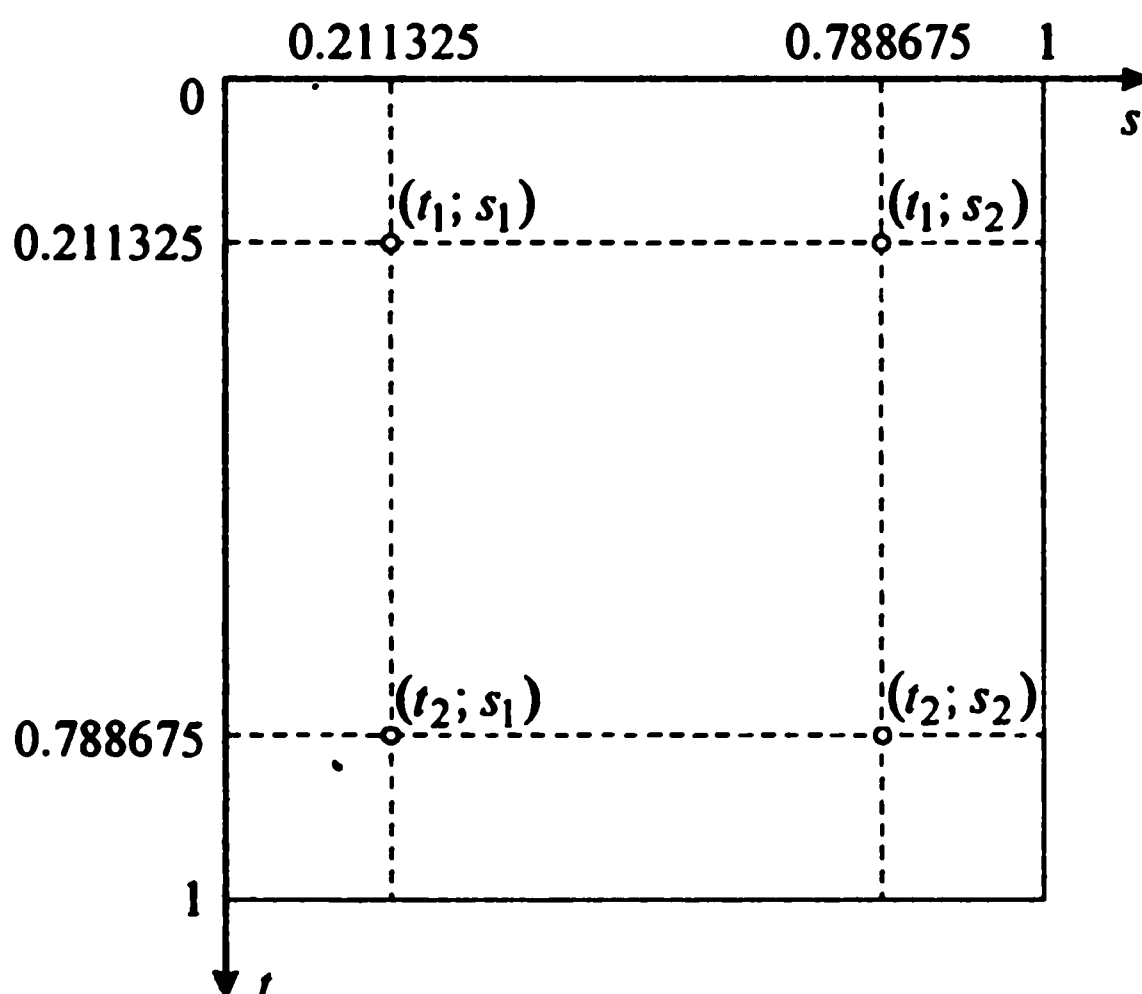


Рис.11.4. Сетка для подсчета каркаса резольвенты на основе квадратурной формулы Гаусса с двумя узлами

В таком случае имеем

$$\hat{Q} \approx \begin{pmatrix} 0 & -0.0962 \\ 0.0962 & 0 \end{pmatrix}, \quad A = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}, \quad U_0 := E = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

и, следовательно, при данном $\lambda = 4$

$$\psi_0 = \lambda A \hat{Q} = \begin{pmatrix} 0 & -0.1925 \\ 0.1925 & 0 \end{pmatrix}.$$

Убедившись в малости $\|\psi_0\|$, что означает выполнение требования (11.66) и, тем более, (11.63), далее продолжаем счет по формулам (11.65):

$$\begin{aligned} U_1 &= \begin{pmatrix} 1 & -0.1925 \\ 0.1925 & 1 \end{pmatrix}, & \psi_1 &= \begin{pmatrix} -0.0370 & 0 \\ 0 & -0.0370 \end{pmatrix}; \\ U_2 &= \begin{pmatrix} 0.9630 & -0.1853 \\ 0.1853 & 0.9630 \end{pmatrix}, & \psi_2 &= \begin{pmatrix} 0.0014 & -0.0000 \\ 0.0000 & 0.0014 \end{pmatrix}; \\ U_3 &= \begin{pmatrix} 0.9643 & -0.1856 \\ 0.1856 & 0.9643 \end{pmatrix}, & \psi_3 &= \begin{pmatrix} 0.0000 & -0.0000 \\ 0.0000 & 0.0000 \end{pmatrix}; \end{aligned}$$

Ясно, что с используемым числом десятичных знаков дальнейшего уточнения значений элементов матриц U_k , приближающих матрицу $(E - \lambda A \hat{Q})^{-1}$, происходить не будет. Значит, искомый каркас резольвенты ядра (11.68) с $\lambda = 4$ на заданной сетке может быть приближенно представлен матрицей

$$R_3 = \hat{Q} U_3 = \begin{pmatrix} 0 & -0.0962 \\ 0.0962 & 0 \end{pmatrix} \begin{pmatrix} 0.9643 & -0.1856 \\ 0.1856 & 0.9643 \end{pmatrix} = \begin{pmatrix} -0.0179 & -0.0928 \\ 0.0928 & -0.0179 \end{pmatrix}.$$

Насколько полученная матрица R_3 близка к истинному каркасу резольвенты на данной сетке, можно выяснить, зная для этого ядра точную резольвенту [19]:

$$R(t, s; 4) = \frac{1}{8} ts(15t - 10ts - 6).$$

Подсчитав ее каркас (с четырьмя знаками после запятой)

$$\bar{R} = \begin{pmatrix} -0.0183 & -0.0937 \\ 0.0867 & -0.0303 \end{pmatrix},$$

находим ошибку приближения R_3 :

$$\bar{R} - R_3 = \begin{pmatrix} -0.0004 & -0.0009 \\ -0.0059 & -0.0124 \end{pmatrix}.$$

Последняя матрица свидетельствует о том, что каркас резольвенты найден, грубо говоря, с точностью до сотых, и эта точность практически

достигается уже на первой итерации, т.е. можно принять

$$\bar{\mathbf{R}} \approx \mathbf{R}_1 = \hat{\mathbf{Q}}\mathbf{U}_1 = \begin{pmatrix} -0.0185 & 0.0962 \\ 0.0962 & -0.0185 \end{pmatrix}$$

с матрицей ошибок

$$\bar{\mathbf{R}} - \mathbf{R}_1 = \begin{pmatrix} 0.0002 & 0.0025 \\ -0.0095 & -0.0118 \end{pmatrix}.$$

Разобранный пример с известной резольвентой показывает, что при построении каркаса резольвенты квадратурно-итерационным методом точность итерационного обращения матрицы $\mathbf{E} - \lambda\mathbf{A}\hat{\mathbf{Q}}$ предопределяется точностью сведения интегральной задачи к матричной и легко регулируется указанием малости величин $\|\psi_k\|$ или $\|\mathbf{U}_k - \mathbf{U}_{k-1}\|$. Ввиду проблем с реальным получением гарантированных оценок погрешностей дискретизации, здесь представляется целесообразным для контроля этих погрешностей квадратур использовать механизм сгущающихся равномерных сеток, общими узлами которых должны служить узлы заданной сетки искомого каркаса.

УПРАЖНЕНИЯ

11.1. По формулам (11.14)–(11.16) решите уравнение с вырожденным ядром (11.19). Сравните полученное таким способом приближенное решение уравнения (11.18) с его точным решением $x(t) \equiv 1$.

11.2. Запишите систему из n уравнений, к которой сводится применение метода конечных сумм на основе квадратурной формулы прямоугольников:

а) для уравнения Гаммерштейна

$$x(t) = \int_a^b Q(t, s) \varphi(s, x(s)) ds, \quad t \in [a, b];$$

б) для уравнения Урысона

$$x(t) = \int_a^b K(t, s, x(s)) ds, \quad t \in [a, b].$$

11.3. Составьте систему уравнений, к которой сводится применение метода конечных сумм на основе квадратурной формулы трапеций с пятью равностоящими узлами для решения *нелинейного уравнения*

Вольтерра

$$x(t) = \int_a^t F(t, s, x(s))ds, \quad t \in [a, b].$$

11.4. Дано уравнение

$$x(t) = \int_1^2 \left(\frac{t}{s^2} - 1 \right) x(s) ds = t^2 + \frac{t}{6} - \frac{7}{3}. \quad (11.69)$$

Найдите его приближенное решение квадратурным методом с тремя узлами, пользуясь:

- а) формулой трапеций;
- б) формулой Гаусса.

В точках $t = 1$, $t = 1.5$ и $t = 2$ сравните полученные результаты с точными значениями решения $x(t)$, найдя его методом вырожденных ядер.

11.5. На уравнениях Вольтерра второго рода (11.39) и первого рода (11.49) примеров 11.3 и 11.4 соответственно, используя те же сетки, исследуйте применение в методе конечных сумм квадратурных формул левых и правых прямоугольников (5.6), (5.7). Сравните результаты между собой и с теми, которые были получены в указанных примерах с помощью других формул численного интегрирования.

11.6. Квадратурно-итерационным методом найдите приближенный каркас резольвенты для ядра уравнения (11.69) (считая $\lambda = 1$) на равномерной 3×3 -сетке узлов квадрата $[1, 2] \times [1, 2]$, применяя:

- а) квадратурную формулу трапеций;
- б) квадратурную формулу Симпсона.

Предварительно оцените, сколько нужно сделать шагов процесса (11.65), чтобы вносимые неточным обращением матриц искажения каркаса резольвенты не превышали 0.001 в случае а) и 0.0001 в случае б).

**ОБРАЗЦЫ ПОСТАНОВОК
ЛАБОРАТОРНЫХ ЗАДАНИЙ**

Лабораторная работа 1

«Интерполяция»

Многократно дифференцируемая функция $y = f(x)$ задана таблицей значений

	x	0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6
Вариант 1	y	1	1.0201	1.0811	1.1855	1.3374	1.5431	1.8107	2.1509	2.5775
Вариант 2	y	0	0.2013	0.4108	0.6367	0.8881	1.1752	1.5095	1.9043	2.3756
...	y	...	...	...	...	...	...	...	...	...

(где последние цифры являются продуктами правильного округления), и заданы контрольные значения аргумента

$\bar{x} = 0.25, \quad \tilde{x} = 0.92 \quad \text{и} \quad \hat{x} = 1.63.$

А) Записать подходящее для приближенного вычисления значений $\bar{y} = f(\bar{x}), \tilde{y} = f(\tilde{x}), \hat{y} = f(\hat{x})$ конкретные интерполяционные многочлены Лагранжа первой и второй степени и получить эти значения.

Б) Составить алгоритм, реализующий схему Эйткена вычисления с максимально возможной точностью значения $y = f(x)$ в произвольной точке x промежутка $[x_0, x_n + (x_n - x_{n-1})]$. Пользуясь этим алгоритмом, вычислить приближенные значения $\bar{y}, \tilde{y}$ и $\hat{y}$.

В) Составить таблицу конечных разностей, записать оптимальные для вычисления $\bar{y}, \tilde{y}$ и $\hat{y}$ конкретные конечноразностные формулы и с их помощью получить эти значения.

Проанализировать результаты выполнения заданий А)–В).

Лабораторная работа 2

«Аппроксимация таблично заданных функций»

Функция $y = f(x)$ задана следующей таблицей значений

	x	10	20	30	40	50	60	70	80
Вариант 1	y	2.5	3.2	3.7	4.0	4.2	4.4	4.6	4.75
Вариант 2	y	4.7	6.7	8.1	9.4	10.6	11.2	12.6	13.4
...	y	...	...	...	...	...	...	...	...

1. Методом наименьших квадратов аппроксимировать $y = f(x)$:

- а) линейной функцией;
- б) многочленами Фурье второй, третьей и четвертой степеней;
- в) функцией вида $a \lg(bx)$;
- г) функцией вида ax^b ;
- д) функцией вида ae^{bx} .

Сравнить величины среднеквадратических отклонений. Пользуясь каждой из найденных функций а)–д), вычислить контрольное приближенное значение $f(35)$.

2. А) Для функции $y = f(x)$ построить интерполяционный кубический сплайн дефекта 1 и с его помощью вычислить приближенно $f(35)$, $f'(35)$ и $\int\limits_{10}^{40} f(x)dx$.

Б) Представить построенный сплайн линейной комбинацией кубических В-сплайнов (конкретизировать вид этих В-сплайнов соответственно рассматриваемому случаю и найти коэффициенты их линейной комбинации).

«Численное интегрирование»

Даны интегралы:

Вариант 1
$$I_1 := \int_{0.4}^2 \frac{1}{x} e^{0.03x} dx,$$

$$I_2 := \int_0^4 \frac{\ln(4+x) - \ln(8)}{(4-x)\sqrt{x}} dx;$$

Вариант 2
$$I_1 := \int_{0.6}^3 \frac{1}{x + \sin 0.5x} dx,$$

$$I_2 := \int_0^2 \frac{\ln(2-x)}{\sqrt{x}\sqrt{2-x}} dx;$$

и т.д.

1. Сколько достаточно взять узлов, чтобы найти значение интеграла I_1 с точностью $\varepsilon_0 = 10^{-3}$:

а) по формулам прямоугольников (средней точки) ?

б) по формулам трапеций ?

Вычислить I_1 по этим формулам с данной точностью.

2. Вычислить I_1 по формуле Симпсона с 9 узлами и по формуле Гаусса с 4 узлами.

3. С точностью $\varepsilon_1 = 10^{-8}$ найти значение интеграла I_1 алгоритмом Ромберга, стараясь минимизировать количество вычислений подынтегральной функции.

Сравнить результаты пунктов 1–3.

4. Подбирая подходящие методы вычисления определенных и несобственных интегралов, с точностью $\varepsilon_2 = 10^{-4}$ найти значение I_2 не менее, чем двумя способами.

Лабораторная работа 4

«Численное дифференцирование»

Бесконечно гладкая функция $y = f(x)$ задана несколькими своими округленными значениями:

	x	0.2	0.3	0.4	0.5	0.6	0.7	0.8
Вариант 1	y	1.3694	1.2661	1.1593	1.0472	0.9273	0.7954	0.6435
Вариант 2	y	0.3948	0.5830	0.7610	0.9272	1.0808	1.2214	1.3494
...	y	...	...	..	...	...	...	...

1. Создать аналогичную таблицу с приближенными значениями функции $f'(x)$, находямыми по формулам:

- а) первого порядка точности,
- б) второго порядка точности,

оставляя в результатах верные цифры и один запасной десятичный знак.

2. Создать таблицу приближенных значений функции $f''(x)$, подсчитываемых по формуле второго порядка точности.

3. Максимально точно, насколько это можно в данных условиях, вычислить значения

$y'(0.25), \quad y'(0.55), \quad y''(0.25) \quad \text{и} \quad y''(0.55).$

Лабораторная работа 5

«Численное решение дифференциальных уравнений первого порядка»

Дано дифференциальное уравнение и начальное условие:

Вариант 1
$$y' = \frac{3x^2 \cos(y^2 - x^3)}{2\sqrt{1 + x^3} \cos(1)}, \qquad y(0) = 1;$$

Вариант 2

$$y' = \frac{2xe^{xy}}{(1+x^2)^{1+x}},$$

$$y(0) = 0;$$

и т.д.

1. Заполнить таблицу

x	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
y											

приближенными значениями решения данной задачи Коши, вычисленными с точностью $\varepsilon = 10^{-8}$ методом Рунге–Кутты с автоматическим выбором шага или методом Кутты–Мерсона (указать окончательный расчетный шаг в каждой точке таблицы).

2. Взяв из таблицы п.1 первые четыре значения решения, продолжить вычисления до точки $x = 1$ с фиксированным шагом $h = 0.1$ методом Милна и предиктор-корректорным методом Адамса четвертого порядка. Подсчитать главные части получаемых при этом на каждом шаге погрешностей.

Лабораторная работа 6

«Численное решение задачи Коши для обыкновенного дифференциального уравнения второго порядка»

Дана задача Коши:

Вариант 1

$$y'' = -\frac{\sqrt{x+y^2}}{4\sqrt{2}x^2},$$

$$y(1) = 1, \quad y'(1) = 0.5;$$

Вариант 2

$$y'' = -\frac{\sqrt{y^2-x}}{2\sqrt{3}x^2},$$

$$y(1) = 2, \quad y'(1) = 1;$$

и т.д.

На отрезке $[1, 2]$ построить таблицу значений ее реше-

ния $y(x)$ с шагом $h = 0.1$ и заданной точностью $\varepsilon = 10^{-6}$, применяя:

а) сведение к системе дифференциальных уравнений первого порядка с последующим численным интегрированием ее методом Рунге–Кутты или Кутты–Мерсона:

б) предиктор-корректорные методы Адамса непосредственно к данной задаче (несколько первых «разгонных» значений можно взять из промежуточных результатов п.а)).

Указать окончательный расчетный шаг, обеспечивающий заданную точность в каждом случае.

Лабораторная работа 7

«Численное решение линейной краевой задачи»

Дана краевая задача:

Вариант 1
$$y'' - \frac{y'}{x} - \frac{3y}{x^2} = \frac{3}{x^2},$$

$$y(0.7) + 0.7y'(0.7) = -1, \quad y(1) = 0;$$

Вариант 2
$$y'' + \frac{y'}{x+2} + \frac{(4x+7)y}{4(x+2)^2} = \frac{1}{\sqrt{x+1}},$$

$$y(2) = 2, \quad y(2.3) = 8.6y'(2.3);$$

и т.д.

На промежутке, определяемом данными краевыми условиями,

1. с точностью $\varepsilon = 10^{-6}$ построить каркас решения $y(x)$ на сетке с шагом $h_0 = 0.1$ конечноразностным методом второго порядка и противопотоковым методом (указать шаг расчетной сетки, при котором обеспечивается эта точность в каждом методе);

2. применить методы Галёркина и коллокации с тремя–четырьмя базисными функциями;

3. решить краевую задачу путём сведения ее к задаче Коши (методом редукции или дифференциальной прогонки).

Результаты пунктов 2, 3 сравнить с результатами пункта 1 (дать сводную таблицу значений приближенных решений на сетке с шагом h_0).

Лабораторная работа № 8

«Квадратурный метод решения интегральных уравнений»

Для уравнений

$$x(t) = \int_0^2 Q(t, s)x(s)ds + f(t) \quad (1)$$

и

$$\int_1^t K(t, s)x(s)ds = F(t) \quad (2)$$

заданы ядра и свободные члены:

Вариант 1	$Q(t, s) = 2 \ln \frac{1+s}{1+t^2},$	$f(t) = t^2 - t + 1;$
	$K(t, s) = t + \sqrt{s},$	$F(t) = 2t\sqrt{t} - t - 1;$
Вариант 2	$Q(t, s) = t + \ln(1+s),$	$f(t) = 1 - \frac{t^2}{t+1};$
	$K(t, s) = \frac{\sqrt{s}-t}{s},$	$F(t) = 3t - 2t\sqrt{t} - 1;$

и т.д.

1. На сетке точек t_i отрезка $[0, 2]$ с шагом сетки $h_1 = 0.5$ построить каркас приближенного решения уравнения (1) с точностью $\varepsilon = 10^{-6}$, пользуясь какой-либо квадратурной формулой замкнутого типа и применяя сгущающиеся расчетные сетки для обеспечения заданной точности. На основе полученного каркаса записать приближенное решение в виде непрерывной функции и с ее помощью вычислить приближенные значения $x\left(\frac{1}{e}\right)$ и $x\left(\frac{\pi}{2}\right)$.

2. Применяя квадратурную формулу прямоугольников на отрезке $[1, 2]$ с шагом $h_2 = 0.2$, найти каркас приближенного решения уравнения (2) с точностью $\varepsilon = 10^{-4}$. Представить полученное дискретное решение интерполяционным многочленом третьей степени, построенным по первым четырем узлам заданной сетки, и, пользуясь этим, вычислить приближенно $x\left(\frac{e^2}{5}\right)$ и $x\left(\frac{\pi^2}{9}\right)$.

ЛИТЕРАТУРА

1. Алберг Дж., Нильсон Э., Уолш Дж. Теория сплайнов и ее приложения. — М.: Мир, 1972.
2. Амосов А.А., Дубинский Ю.А., Копчёнова Н.В. Вычислительные методы для инженеров. — М.: Высшая школа, 1994.
3. Андреев В.Б., Руховец Л.А. Проекционные методы // Математика, кибернетика, №11.— М.: Знание, 1986.
4. Арушанян О.Б., Залеткин С.В. Численное решение обыкновенных дифференциальных уравнений на Фортране. — М.: Изд-во МГУ, 1990.
5. Ахиезер Н.И. Лекции по теории аппроксимации. — М.: Наука, 1965.
6. Бабенко К.И. Основы численного анализа. — М: Наука, 1986.
7. Бабушка И., Витасек Э., Прагер М. Численные процессы решения дифференциальных уравнений. — М.: Мир, 1969.
8. Бахвалов Н.С. Численные методы. — М.: Наука, 1973.
9. Бахвалов Н.С., Жидков Н.П., Кобельков Г.М. Численные методы. — М.: Наука, 1987.
10. Бахвалов Н.С., Лапин А.В., Чижонков Е.В. Численные методы в задачах и упражнениях. — М.: Высшая школа, 2000.
11. Березин И.С., Жидков Н.П. Методы вычислений. Т.1. — М.: Физматгиз, 1962.
12. Березин И.С., Жидков Н.П. Методы вычислений. Т.2. — М.: Физматгиз, 1962.
13. Боглаев Ю.П. Вычислительная математика и программирование. — М.: Высшая школа, 1990.
14. Бор К. Практическое руководство по сплайнам. — М.: Радио и связь, 1985.
15. Вапник В.Н. Восстановление зависимостей по эмпирическим данным. — М.: Наука, 1979.
16. Варга Р. Функциональный анализ и теория аппроксимации в численном анализе. — М.: Мир, 1974.

17. *Васин В.В., Агеев А.Л.* Катастрофы и парадоксы при решении неустойчивых задач на ЭВМ // Математика, кибернетика, №10. — М.: Знание, 1991.
18. *Вержбицкий В.М.* Численные методы (линейная алгебра и нелинейные уравнения). — М.: Высшая школа, 2000.
19. *Верлань А.Ф., Сизиков В.С.* Интегральные уравнения: методы, алгоритмы, программы. — Киев: Наукова думка, 1986.
20. *Воеводин В.В., Кузнецов Ю.А.* Матрицы и вычисления. — М.: Наука, 1984.
21. *Волков Б.А.* Численные методы. — М.: Наука, 1979.
22. *Вуколов Э.А., Ефимов А.В. и др.* Сборник задач по математике для втузов. Ч.4. Методы оптимизации. Уравнения в частных производных. Интегральные уравнения. — М.: Наука, 1990.
23. *Вулих Б.З.* Введение в функциональный анализ. — М.: Наука, 1967.
24. *Гельфонд А.О.* Исчисление конечных разностей. — М.: Физматгиз, 1959.
25. *Гилл Ф., Мюррей У., Райт М.* Практическая оптимизация. — М.: Мир, 1985.
26. *Гончаров В.Л.* Теория интерполирования и приближения функций. — М. — Л.: ГТТИ, 1934.
27. *Гутер Р.С., Овчинский Б.В.* Элементы численного анализа и математической обработки результатов опыта. — М.: Наука, 1970.
28. *Двайт Г.Б.* Таблицы интегралов и другие математические формулы. — М.: Наука, 1966.
29. *Деккер К., Вервер Я.* Устойчивость методов Рунге-Кутты для жестких нелинейных дифференциальных уравнений. — М.: Мир, 1988.
30. *Демидович Б.П., Марон И.А.* Основы вычислительной математики. — М.: Наука, 1970.
31. *Демидович Б.П., Марон И.А., Шувалова Э.З.* Численные методы анализа. — М.: Наука, 1967.
32. *Демьянов В.Ф., Малоземов В.Н.* Введение в минимакс. — М.: Наука, 1972.

33. *Дробышев В.И., Дымников В.П., Ривин Г.С.* Задачи по вычислительной математике. — М.: Наука, 1980.
34. *Дьяченко В.Ф.* Основные понятия вычислительной математики. — М.: — Наука, 1972.
35. *Журкин И.Г., Нейман Ю.М.* Методы вычислений в геодезии. — М.: Недра, 1988.
36. *Забрейко П.П., Кошелев А.И. и др.* Интегральные уравнения. — М.: Наука, 1968.
37. *Иванов В.В.* Методы вычислений на ЭВМ: Справочное пособие. — Киев: Наукова думка, 1986.
38. *Иванова В.М., Калинина В.Н. и др.* Математическая статистика. — М.: Высшая школа, 1975.
39. *Калиткин Н.Н.* Численные методы. — М.: Наука, 1978.
40. *Камке Э.* Справочник по обыкновенным дифференциальным уравнениям. — М.: Наука, 1976.
41. *Канторович Л.В., Акилов Г.П.* Функциональный анализ. — М.: Наука, 1977.
42. *Канторович Л.В., Крылов В.И.* Приближенные методы высшего анализа. — М.-Л.: Физматгиз, 1962.
43. *Каханер Д., Моулер К., Нэш С.* Численные методы и программное обеспечение. — М.: Мир, 1998.
44. *Коваленко И.Н., Филиппова А.А.* Теория вероятностей и математическая статистика. — М.: Высшая школа, 1973.
45. *Копченкова Н.В., Марон И.А.* Вычислительная математика в примерах и задачах. — М.: Наука, 1972.
46. *Корн Г., Корн Т.* Справочник по математике для научных работников и инженеров. — М.: Наука, 1968.
47. *Корнейчук Н.П.* Сплайны в теории приближений. — М.: Наука, 1984.
48. *Косарев В.И.* 12 лекций по вычислительной математике. — М.: Изд-во МФТИ, 1995.
49. *Кочкиков И.В., Курамшина Г.М., Ягола А.Г.* Численные методы в колебательной спектроскопии // Математика, кибернетика, №1. — М.: Знание, 1989.

50. *Краскевич В.Е., Зеленский К.Х., Гречко В.И.* Численные методы в инженерных исследованиях. — Киев: Вища школа, 1986.
51. *Краснов М.Л.* Интегральные уравнения. — М.: Наука, 1975.
52. *Краснов М.Л., Киселев А.И., Макаренко Г.И.* Интегральные уравнения. Задачи и упражнения. — М.: Наука, 1968.
53. Приближенное решение операторных уравнений/*Красносельский М.А., Вайникко Г.М., Забрейко П.П. и др.* — М.: Наука, 1969.
54. *Крылов В.И., Бобков В.В., Монастырный П.И.* Вычислительные методы. Т.1. — М.: Наука, 1976.
55. *Крылов В.И., Бобков В.В., Монастырный П.И.* Вычислительные методы. Т.2. — М.: Наука, 1977.
56. *Крылов В.И., Бобков В.В., Монастырный П.И.* Начала теории вычислительных методов. Интерполирование и интегрирование. — Минск.: Наука и техника, 1983.
57. *Крылов В.И., Шульгина Л.Т.* Справочная книга по численному интегрированию. — М.: Наука, 1966.
58. *Курант Р.* Курс дифференциального и интегрального исчисления. Т.1. — М.: Наука, 1967.
59. *Лизоркин П.И.* Курс дифференциальных и интегральных уравнений с дополнительными главами анализа. — М.: Наука, 1981.
60. *Ловитт У.В.* Линейные интегральные уравнения. — М.: ГИТТЛ, 1957.
61. *Люстерник Л.А., Червоненкис О.А., Янпольский А.Р.* Математический анализ. Вычисление элементарных функций. — М.: Физматгиз, 1963.
62. *Макаров В.Л., Хлобыстов В.В.* Сплайн-аппроксимация функций. — М.: Высшая школа, 1983.
63. *Марчук Г.И.* Методы вычислительной математики. — М.: Наука, 1977.
64. *Михлин С.Г., Смолицкий Х.Л.* Приближенные методы решения дифференциальных и интегральных уравнений. — М.: Наука, 1965.

65. Сборник задач по методам вычислений/ Под ред. *П.И.Монастырного* — М.: Наука, 1994.
66. *Молчанов И.Н.* Машинные методы решения прикладных задач. Дифференциальные уравнения. — Киев: Наукова думка, 1988.
67. *Мысовских И.П.* Лекции по методам вычислений. — М.: Физматгиз, 1962.
68. *Мышкис А.Д.* Математика для втузов. Специальные курсы. — М.: Наука, 1971.
69. *Никифоров А.Ф., Суслов С.К.* Классические ортогональные полиномы // Математика, кибернетика, №12. — М.: Знание, 1985.
70. *Никольский С.М.* Квадратурные формулы. — М.: Наука, 1988.
71. *Носач В.В.* Решение задач аппроксимации с помощью персональных компьютеров. — М.: МИКАП, 1994.
72. *Норри Д., де Фриз Ж.* Введение в метод конечных элементов. — М.: Мир, 1981.
73. *Ортега Дж., Пул У.* Введение в численные методы решения дифференциальных уравнений. — М.: Наука, 1986.
74. *Плис А.И., Сливина Н.А.* Лабораторный практикум по высшей математике. — М.: Высшая школа, 1994.
75. *Попов Б.А., Теслер Г.С.* Приближение функций для технических приложений. — Киев: Наукова думка, 1980.
76. *Пугачев В.С.* Лекции по функциональному анализу. — М.: Изд-во МАИ, 1996.
77. *Ракитин В.И., Первушин В.Е.* Практическое руководство по методам вычислений с приложением программ для персональных компьютеров. — М.: Высшая школа, 1998.
78. *Ремез Е.Я.* Основы численных методов чебышевского приближения. — Киев: Наукова думка, 1969.
79. *Рябенский В.С.* Введение в вычислительную математику. — М.: Наука, 1994.
80. *Самарский А.А.* Введение в численные методы. — М.: Наука, 1987.

81. Самарский А.А., Гулин А.В. Численные методы. — М.: Наука, 1989.
82. Смирнов В.И. Курс высшей математики. Т.2. — М.: Наука, 1967.
83. Степанов В.В. Курс дифференциальных уравнений. Изд.6. — М.: ГИТТЛ, 1953.
84. Стечкин С.Б., Субботин Ю.Н. Сплайны в вычислительной математике. — М.: Наука, 1976.
85. Суетин П.К. Классические ортогональные многочлены. — М.: Наука, 1976.
86. Треногин В.А. Функциональный анализ. — М.: Наука, 1980.
87. Турчак Л.И. Основы численных методов. — М.: Наука, 1987.
88. Форсайт Дж., Молер К. Численное решение систем линейных алгебраических уравнений. — М.: Мир, 1969.
89. Хайрер Э., Нёрсетт С., Ваннер Г. Решение обыкновенных дифференциальных уравнений. Нежесткие задачи. — М.: Мир, 1990.
90. Хемминг Р.В. Численные методы. — М.: Наука, 1968.
91. Холл Дж., Уатт Дж. Современные численные методы решения обыкновенных дифференциальных уравнений. — М.: Мир, 1979.
92. Цимринг Ш.Е. Специальные функции и определенные интегралы. Алгоритмы. Программы для микрокалькуляторов: Справочник. — М.: Радио и связь, 1988.
93. Цлаф Л.Я. Вариационное исчисление и интегральные уравнения. — М.: Наука, 1970.
94. Штеттер Х. Анализ методов дискретизации для обыкновенных дифференциальных уравнений. — М.: Мир, 1978.
95. Математическая энциклопедия. ТТ. 1–5. — М.: Советская энциклопедия, 1977.
96. Математический энциклопедический словарь. — М.: Советская энциклопедия, 1988.
97. Словарь иностранных слов. — М.: Русский язык, 1989.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

- Аддитивное выделение особенностей 184
Алгебраический порядок точности 169, 248
Алгоритм прямоугольников-трапеций 165
— Ромберга 166
Аналоговая вычислительная машина 211
Апостериорный контроль глобальной погрешности 232
Аппроксимация 8
— кусочно-полиномиальная 9, 111
— полиномиальная 9

Базисный многочлен Лагранжа 13

Вес (весовой коэффициент) квадратурной формулы 167
Весовая функция 175
Вторая интерполяционная формула Гаусса 38
— — — Ньютона 36
Второй интерполяционный многочлен Ньютона 35

Глобальная погрешность 151

Дефект сплайна 117
Диаграмма Фрезера 43

Жесткая система 284
— — на интервале 285

Задача Абеля 333
— интерполяции 12
— краевая 294, 295

Задача начальная (Коши) 210
— некорректная 336
— обратного интерполирования 51
Звено сплайна 125

Интегральное уравнение 212, 333
Интегральное уравнение линейное 334
— — нелинейное Вольтерра 362
Интерполирование 11
Интерполяционная квадратурная формула 177
— схема Эйткена 22
— формула Бесселя 38
— — Ньютона для неравноотстоящих узлов 49
— — Стирлинга 38
Интерполяция 11

Каркас приближенного решения 264, 342
— резольвенты 357
Квадратура Лобатто 174
— механическая 146
Квадратурная формула 146
— — Гаусса 170
— — Гаусса–Кристоффеля 175
— — замкнутого типа 174
— — интерполяционного типа 177
— — Лагерра 179
— — левых прямоугольников 147
— — Маркова 174
— — Мелера 178
— — Ньютона–Котеса 152
— — открытого типа 174
— — правых прямоугольников 147

- Квадратурная формула средних прямоугольников 148
 — — типа Гаусса 175
 — — Чебышева 169
 — — Чебышева–Лагерра 179
 — — Эйлера 175
 — — — составная 175
 — — Эрмита 178, 179
 Конечноразностная формула численного дифференцирования 190
 Конечные разности 26
 — — многочлена 28
 — — практически постоянные 32
 Конечный элемент 330
 Корневое условие 278
 Коэффициент Котеса 152
 — Фурье 96
 Кратность узла 56
 Критерий согласия 9
 Критерий Чебышева 74

 Линейное интегральное уравнение второго рода Вольтерра 335
 — — — — — Фредгольма 335
 — — — первого рода Вольтерра 336
 — — — — — Фредгольма 336
 — — — третьего рода 337
 — разностное уравнение с постоянными коэффициентами 271
 Линейный фильтр 116
 Линейных многошаговых методов общий вид 248

 Матрица Гильберта 97
 Мера аппроксимации 264
 — обусловленности 98
 Метод абсолютно устойчивый 275
 — A -устойчивый 286
 — $A(\alpha)$ -устойчивый 290
 — вложенных форм 231
 — Галёркина 315

 Метод графический 211
 — двухшаговый 220
 — двухэтапный 225
 — дифференциальной прогонки 301
 — дифференцирования назад 291
 — замены ядра на вырожденное 340
 — квадратурный (конечных сумм) 340, 341, 345
 — коллокации (совпадений, интерполяционный) 309
 — конечных разностей (МКР) 302
 — — элементов (МКЭ) 320
 — Коуэлла четвертого порядка 252
 — Кутты–Мерсона 231
 — Милна 244
 — — второго порядка 220
 — моментов 340
 — наименьших квадратов (МНК) 86
 — Нистрёма второго порядка 220
 — обратной линейной интерполяции 52
 — Пикара 213
 — последовательных приближений 213, 340
 — пристрелки (стрельбы) 298
 — прогноза и коррекции (предсказания и уточнения) 241
 — прогонки 122
 — проекционно-разностный (проекционно-сеточный) 320
 — проекционный 314
 — противопотоковый (upwind) 308
 — редукции 298
 — Рунге 319
 — Рунге–Кутты 223
 — — — четвертого порядка 226
 — — — — — пятиэтапный 231
 — с забеганием вперед 252
 — средней точки 225
 — степенных рядов 211
 — трапеций 218, 240, 250

Метод условно устойчивый 274
 — Хойна (Хьюна) 219, 225
 — четырехэтапный 226
 — чисто неявный 291
 — Штёрмера 260
 — Эйлера (ломаных) 215, 238, 250
 — — исправленный 222
 — — неявный (обратный) 218, 240, 250
 — — с пересчетом 219
 — — уточненный 220
 — — явно-неявный 241
 Многочлен Лагерра 102
 — Лагранжа интерполяционный 13
 — Лежандра 100
 — наилучшего равномерного приближения 74
 — Чебышева 64
 — — второго рода 101
 — — нормированный 66
 — — первого рода 101
 — — смещенный 69
 — Эрмита 102
 — — интерполяционный 56
 — Якоби 101
 Многошаговый метод Адамса 236
 Модельное уравнение 273
 Модификация Ингланда 232
 — Фельберга 232
 Мультипликативное выделение особенностей 184

 Наилучшая равномерная оценка погрешности интерполяции 71
 Некорректность задачи численного дифференцирования 207
 Норма чебышевская 73
 Нормальная система МНК 94
 Нормальное псевдорешение 86

 Область устойчивости 286

Обобщенный многочлен 93
 — — наилучшего среднеквадратического приближения 93
 — — Фурье 96
 Общая формула прямоугольников 146
 Оператор восполнения 263
 — сноса 263
 Определитель Фредгольма 338
 Оптимальный шаг численного дифференцирования 205
 Ординатный вид формулы 237
 Ортогональность с весом 101
 Осреднение по трем точкам 116
 Остаточный член интерполяционной формулы 16
 — — — Эрмита 61
 — — простейшей формулы прямоугольников 149
 — — — Симпсона 154
 — — — трапеций 153
 — — формулы Симпсона 160
 — — — трапеций 159
 Отрезок тригонометрического ряда Фурье 96
 Ошибка глобальная 216
 — локальная (шаговая) 216

 Первая интерполяционная формула Гаусса 37
 — — — Ньютона 34
 Первый интерполяционный многочлен Ньютона 33
 Погрешность шаговая 223
 Поправка Ричардсона 163, 229
 — — обобщенная 165
 — шаговая 225
 Порядок аппроксимации 264
 Предиктор-корректорный метод 241
 — — — Адамса 241
 Приближение 8

- Приближение наилучшее средне-
 квадратическое 93, 106
 — чебышевское 74
 Приближенное решение 264
 Принцип Рунге 164, 228
 Простейшая квадратурная форму-
 ла Симпсона 154
 — — — трапеций 153
 Пространство гильбертово 314
- Разделенные разности 45
 Разностная схема 266
 Разностное уравнение 266
 — — однородное 271
 Разностные отношения 46
 Резольвента (разрешающее ядро) 355
 Решение разностного уравнения
 общее 271
 — — — фундаментальное 271
 — — — частное 271
 Ряд Неймана 356
- Сглаживание 114
 Семейство методов Рунге–Кутты
 второго порядка 224
 — многошаговых методов Адамса
 236
 Сетка 26, 266, 303
 Система фундаментальных реше-
 ний 272
 Среднеквадратическая ошибка ин-
 тегральная 93
 — — точечная 93
 Сплайн 117
 — базисный (*B*-сплайн) 133
 — — квадратичный 136
 — — кубический 137
 — — линейный 135
 — — нулевой степени 134
 — естественный (чертежный) 118
 — интерполяционный 117
- Сплайн квадратичный 125
 — кубический дефекта 1 118
 — локальный 139
 — сглаживающий 124
 — эрмитов 139
 Сходимость каракасов прибли-
 женных решений 265
 — приближенных решений 264
- Таблица конечных разностей 29
 Теорема Вейерштрасса 73
 — Чебышева 67, 74
 Точка расчетная 214
 — чебышевского альтернанса 75
- Узел 10
 — базовый 34
 — интерполяции 11
 — — чебышевский 71
 — квадратичного сплайна 125
 — квадратурной формулы 167
 — кратный 56
 — сетки 266
 — сплайна 117
 Узлы равноотстоящие 26
 Уравнение Гаммерштейна 362
 — диффузии одномерное 318
 — операторное 314
 — разностное 266
 — теплопроводности одномерное
 стационарное 318
 — Урысона 362
 — характеристическое 271
 Условие корней (условие α) 278
 — краевое 295
 Условия интерполяции 12
 — согласованности параметров 249
 — эрмитовой интерполяции 56
 Усовершенствованный метод Эй-
 лера–Коши 219

**Устойчивость конечноразностной
схемы 306**

— по Дальквисту 278

— — — строгая (сильная) 278

Формула Грегори 157

— интерполирования на середину
39

— интерполяционная 15

— квадратичной интерполяции 14

— линейной интерполяции 14

— Милна вторая (уточнения) 246

— — первая (предсказания) 245

— прямоугольников 148

— Родрига 100

— Симпсона 160

— средней точки 148

— трапеций 158

— численного интегрирования 146

— Штёрмера (Адамса–Штёрмера)
260

Функция базисная 308, 315

— интерполирующая 11

— координатная 315

— кусочно-квадратичная 112

— кусочно-линейная 112

— сеточная 11, 266, 303

— табличная 11

Функция финитная 320

**Центральные интерполяционные
формулы 36**

— разности 36

Число жесткости 284

Шаг аппроксимации 264

— расчетный 214

— сетки 26

Шаговость метода 239

Шум округлений 31

**Экономизация степенного разло-
жения 82**

— — ряда 79

Экстраполяционный метод 239

— — Адамса–Башфорта 237

Экстраполяция 14

**Явный четырехшаговый метод
Адамса 258**

Ядро вырожденное 338

— интегрального уравнения 335

— итерированное 356

— мультипликативное 338

— разрешающее (резольвента) 355

Учебное издание

Вержбицкий Валентин Михайлович

ЧИСЛЕННЫЕ МЕТОДЫ
(математический анализ и обыкновенные
дифференциальные уравнения)

Редактор Л.В. Честная

Художественный редактор Ю.Э. Иванова

Оригинал-макет выполнен Ю.В. Гавриным

ЛР № 010146 от 25.12.96. Изд. № ФМ-212. Сдано в набор
и подп. в печать 22.01.2001. Формат 60х88 $\frac{1}{16}$. Бум. газетн.
Гарнитура «Таймс». Печать офсетная. Объем: 23,52 усл. печ. л.,
23,52 усл. кр.-отт., 20,66 уч.-изд. л. Тираж 8000 экз. Заказ № 295

ГУП «Издательство «Высшая школа», 127994, Москва, ГСП-4,
Неглинная ул., д. 29/14

Факс: 200-03-01, 200-06-87. E-mail: V-Shkola@g23.relcom.ru
[http: // www.v-shkola.ru](http://www.v-shkola.ru)

Отпечатано в ГУП ИПК «Ульяновский Дом печати»,
432980, г. Ульяновск, ул. Гончарова, 14

Вержбицкий В.М.

В31 Численные методы (математический анализ и обыкновенные дифференциальные уравнения): Учеб. пособие для вузов. — М.: Высш. шк., 2001. — 382 с.:ил.

ISBN 5-06-003982-X

Рассматриваются вопросы приближения функций интерполяционными многочленами, обобщенными многочленами Фурье и сплайнами. На основе интерполирования выводятся различные формулы численного дифференцирования и интегрирования. Изучаются одношаговые и многошаговые методы решения начальных задач для обыкновенных дифференциальных уравнений, исследуется их численная устойчивость; для краевых задач даются как приближенно-аналитические, так и собственно численные методы. Показываются способы построения каркасов решений линейных интегральных уравнений и их резольвент. Изложение теории сопровождается демонстрационными примерами, таблицами, рисунками; каждая глава завершается упражнениями. В приложении можно найти образцы постановок лабораторных заданий.

Предлагаемое издание продолжает книгу автора «Численные методы (линейная алгебра и нелинейные уравнения)». — М.: Высш. шк., 2000, но может использоваться независимо от нее всеми, кто интересуется вычислительной математикой.

**УДК 519.6
ББК 22.19**